DOCUMENT RESUME

ED 352 378 TM 019 275

TITLE	The OECD International Education Indicators: A Framework for Analysis.
INSTITUTION	Organisation for Economic Cooperation and Development, Paris (France). Centre for Educational Research and Innovation.
REPORT NO	ISBN-92-64-13726-2
PUB DATE	92
NOTE	116p.
AVAILABLE FROM	Publications Service, Organisation for Economic Co-Operation and Development, 2, Rue Andre-Pascal, 75775 Paris CEDEX 16, France.
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Agency Role; Classification; *Comparative Analysis; Concept Formation; Data Analysis; Decentralization; Decision Making; *Educational Research; Elementary Secondary Education; Foreign Countries; *International Studies; Models; *Research Projects
IDENTIFIERS	*Centre for Educ Research and Innovation (France); *Educational Indicators

ABSTRACT

This volume contains some of the many studies undertaken during the initial phase of activity by the Centre for Educational Research and Innovation (CERI) of the Organisation for Economic Cooperation and Development (OECD) concerning international education indicators. The studies trace the preparation of a set of international indicators as they illuminate the complexity of such indicators and the wariness with which decision makers approach indicators. The first inquiries were aimed at circumscribing the field of intervention of the CERI and they illustrate that the most complex problem is not the calculation of valid indicators but the classification of concepts. After an Introduction, "What Are International Education Indicators For?" (N. Bottani and H. J. Walberg), the following articles are included: (1) "The Functions and Limitations of International Education Indicators" (D. Nuttall); (2) "Conceptual Models in Use for Education Indicators" (M. van Herpen); (3) "Process Indicators of School Functioning" (J. Scheerens); (4) "Do Common Values Produce Common Indicators?" (A. Ruby); (5) "Statistical and Indicators Work in the OECD Directorate for Science, Technology and Industry" (G. Westholm); (6) "Standardizing Data in Decentralised Educational Data Systems" (R. W. Selden); and (7) "What Can Be Learned from All This?" (I. Delfau and R. W. Selden). (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Centre for Educational Research and Innovation

The OECD International Education Indicators

A Framework for Analysis

ED 362878

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

J. R. WEST

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM019275

BEST COPY AVAILABLE

Centre for Educational Research and Innovation

The OECD International Education Indicators

A Framework for Analysis

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

- to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971) and New Zealand (29th May 1973). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention). Yugoslavia has a special status at OECD (agreement of 28th October 1961).

The Centre for Educational Research and Innovation was created in June 1968 by the Council of the Organisation for Economic Co-operation and Development.

The main objectives of the Centre are as follows:

- *to promote and support the development of research activities in education and undertake such research activities where appropriate;*
- *to promote and support pilot experiments with a view to introducing and testing innovations in the educational system;*
- *to promote the development of co-operation between Member countries in the field of educational research and innovation.*

The Centre functions within the Organisation for Economic Co-operation and Development in accordance with the decisions of the Council of the Organisation, under the authority of the Secretary-General. It is supervised by a Governing Board composed of one national expert in its field of competence from each of the countries participating in its programme of work.

Publié en français sous le titre :

L'OCDE ET LES INDICATEURS INTERNATIONAUX
DE L'ENSEIGNEMENT
Un cadre d'analyse

© OECD 1992
Applications for permission to reproduce or translate all or part of this publication should be made to:
Head of Publications Service, OECD
2, rue André-Pascal, 75775 PARIS CEDEX 16, France

FOREWORD

This volume contains some of the many studies undertaken during the initial phase of the CERI activity on international education indicators.

These have both historical and conceptual value. Historically, they trace the preparation of a set of international education indicators. Although this is not an entirely new activity – a first attempt at developing indicators was made at the OECD in the early 1970s – it was given a fresh impetus in the late 1980s thanks to the changed political context that has greatly influenced the debate on education and educational policy reforms. The studies reveal how wary decision-makers are of indicators while being attracted by their potential ability. They also illuminate the complexity of the very concept of indicators.

The first phase of the CERI activity, the INES Project, which lasted from June 1988 to December 1989, examined the feasibility of developing and constructing comparable indicators on education systems in OECD countries, in order to permit an assessment of quality and ways to improve it. The conclusion was that it was not only possible, but also desirable to assemble at the international level a set of indicators in a field as critical and controverted as education, provided that the approach be thoroughly systematic and the expectations modest. No one should underestimate the technical problems to be faced, and the nature of the theoretical and ideological challenge.

The first enquiries were aimed at circumscribing the field of intervention of the CERI activity. They should initiate a stimulating and critical debate as the precondition of addressing educational and political reality, extending our knowledge and producing useful analytical tools for promoting quality in education.

This volume is published on the responsibility of the OECD Secretary-General.

ALSO AVAILABLE

Education in OECD Countries

A Compendium of Statistical Information 1987-1988 (1991)
(91 90 06 1) ISBN 92-64-13425-5 FF140 £17.00 US$30.00 DM55

Main Science and Technology Indicators
(94 00 00 3)
Subscription FF 170 £ 20.00 US$ 36.00 DM 67

OECD Employment Outlook

1991 (1991)
(81 91 02 1) ISBN 92-64-13540-5 FF200 £26.00 US$48.00 DM78

The State of the Environment. With Free Supplement: Environmental Indicators. A Preliminary Set (1991)
(97 91 01 1) ISBN 92-64-13442-5 FF180 £22.00 US$38.00 DM70

Prices charged at the OECD Bookshop.

THE OECD CATALOGUE OF PUBLICATIONS and supplements will be sent free of charge on request addressed either to OECD Publications Service, or to the OECD Distributor in your country.

TABLE OF CONTENTS

Introduction

WHAT ARE INTERNATIONAL EDUCATION INDICATORS FOR?
by Norberto Bottani and Herbert J. Walberg 7

Chapter 1

THE FUNCTIONS AND LIMITATIONS OF INTERNATIONAL EDUCATION INDICATORS
by Desmond Nuttall 13

Chapter 2

CONCEPTUAL MODELS IN USE FOR EDUCATION INDICATORS
by Max van Herpen 25

Chapter 3

PROCESS INDICATORS OF SCHOOL FUNCTIONING
by Jaap Scheerens 53

Chapter 4

DO COMMON VALUES PRODUCE COMMON INDICATORS?
by Alan Ruby 77

Chapter 5

STATISTICAL AND INDICATORS WORK IN THE OECD DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY
by Gunnar Westholm 83

Chapter 6

STANDARDIZING DATA IN DECENTRALISED EDUCATIONAL DATA SYSTEMS
by Ramsay W. Selden 107

Chapter 7

WHAT CAN BE LEARNED FROM ALL THIS?
by Isabelle Delfau and Ramsay W. Selden 115

Introduction

WHAT ARE INTERNATIONAL EDUCATION INDICATORS FOR?

by

Norberto Bottani
OECD/CERI Secretariat

and

Herbert J. Walberg
Professor, University of Illinois
United States

I. THE CURRENT INTEREST IN INDICATORS

Under the initiative and invitation of the US Department of Education, and jointly organised with the OECD Secretariat, an international conference on education indicators took place in November 1987 in Washington, D.C. It was attended by national representatives of 22 Member countries, as well as invited experts and observers from the United States and other OECD countries. The need for information and bench-marks that might allow comparisons across countries and indications of how well education is functioning in each country was seen as an integral component of the improvement of the quality of education. Though the conference revealed national differences of priority concerning how schooling is to be improved, the importance of that broad goal and the need for information to support it were echoed repeatedly by country representatives.

Nevertheless, the peculiarity of practices and arrangements in each country implies that cross-national comparisons, even through relatively simple indicators such as enrolments or advancement rates, could not be undertaken lightly. Sometimes goals and cultures are so specific to each country to cause difficulties of measurement and comparison.

Despite such problems, policy-makers would in any event use information and make comparisons using available indicators, irrespective of the caveats provided by the research community. Therefore, it seemed wise to invest considerable effort to produce extensive, reliable, and valid indicators. To be useful, however, the research would also have to be closely tailored to the needs of policy-making. Education, as a human enterprise, might then respond to the information, whether positive or negative. An international organisation such as the OECD is strategically placed to span the divide

between responding to immediate demands for knowledge and the long-term policy-relevant questions that are not amenable to routine measurement and response.

The Washington Conference invited the OECD to carry out developmental work on a comparative set of education indicators. This initiative was confirmed by the International Conference on the Evaluation of Education Systems held in Poitiers in March 1988 at the invitation of the French Ministry of Education.

Both conferences confirmed the urgent need for better and more comprehensive information about the outcomes of education. The current educational debate is characterised by growing concern with the qualitative aspects of education systems, over and above the traditional management questions that arose as a result of the massive expansion of education systems in the post-war period.

The concern about discrepancies and disparities in education systems is not only brought about by exceptional growth. It emanates also from new attitudes that have influenced and changed educational expectations. At the same time, evaluation mechanisms more sensitive to qualitative aspects of public service institutions are being developed. They have had implications for educational policy which is being developed in a very different context from that of ten years ago and is forcing education authorities to rethink the issues, analyse the problems from different angles, and, when necessary, to adopt new objectives.

II. MONITORING AND COMPARING THE PERFORMANCE OF EDUCATION SYSTEMS

Education remains a top priority in social and economic policy for OECD countries. The intellectual demands of work are growing and the sum of human knowledge is increasing. Economic competition and co-operation open new opportunities for countries to enhance the effectiveness and productivity of their education systems.

These developments are causing political and educational leaders to ask if the capacities of youth in their countries are sufficient in both quantitative and qualitative respects. They want to know what degree of educational progress is being made and how their education systems compare with others. This information is required for national reporting, to guide school improvement, to help in priority setting, and to meet increased demands for accountability.

This information cannot be best employed for policy purposes unless education systems are systematically monitored over time and unless the information obtained about the systems can be compared internationally. Indicators should therefore be developed at the significant levels of education systems – the school, the local level, the national or system level and the international level. This is why the CERI International Indicators Project, that brings together individual efforts in setting up national information systems with international comparability, has been characterised by keen national interest and international collaboration.

Experience over the past few decades has shown that when several countries collaborate on evaluation schemes and pursue shared scientific standards, they increase the value and validity of their joint and individual efforts. Countries, of course, need time to reach agreements and to arrive at a common set of scientific meanings and standards, to

test the preliminary systems in close collaboration, to refine the original instruments and methods, and finally to derive an established and enduring set of procedures.

For this reason the concrete specification of international education indicators (Phase II) was scheduled to take place after an initial and exploratory phase. The exploratory stage (Phase I) was intended to demonstrate the feasibility of constructing internationally comparable education indicators.

Specifically and briefly, the first phase of the project aimed to:

- develop a preliminary exploration of a set of key indicators provided by the participating countries;
- provide a forum for exchange about the development of indicators and their use in educational policy and management; and,
- contribute to evaluation methodology, indicator development, and understanding of the use, relations, and limitations of education indicators.

Phase I has confirmed the need for an indicator system and demonstrated its feasibility. It has also confirmed that the implementation of an international information base for more appropriate comparisons requires a suitable organisation. The complexity of the issues involved and the political and scientific relevance of the outcomes for policy-makers demand high levels of reliability. To be useful, indicators must be not only valid and accurate but also comparable. This result can only be achieved by standardizing data, a task which must confront the differences and the autonomy of the various education systems.

The most complex problem is not the calculation of valid indicators but the classification of concepts. Once appropriate data have been specified and collected, the calculations can follow. One of the main items under discussion therefore is the construction of an appropriate, effective, international information system within existing policy-making institutions. The introduction of a new information system will have to be combined with organisational changes in order to construct indicators on the quality of the education systems. This is a main challenge for international co-operation in the field of education in the next several years. But, as said in the first OECD publication on education indicators, "if either the goals or desired states of the system are undefined or unclear, or the means to attain them are unknown, then no information is useful and anything or nothing will serve as an indicator" (*Indicators of Performance of Educational Systems*, OECD, Paris, 1973, p. 15).

The aim is not to create international rankings but rather to provide national policy-makers with a broader understanding of factors influencing the quality of education and to permit them to explore a wider range of policy options. In theory, it is desirable to develop a wide range of indicators that accommodates the breadth of existing educational objectives within countries and that allows countries to choose those measures that best match their goals and priorities. In practice, the cost and difficulty of establishing a wide range of indicators adds weight to the argument for concentrating on a limited set of indicators that are acknowledged by all to be significant and to explore how these can be interpreted for divergent national priorities. For these reasons, a major focus of discussion during Phase I has been the appropriate balance of parsimony and comprehensiveness.

III. FIRST STEPS

When indicators describe generic features of education systems that are defined and measured consistently, they can be used to compare conditions both within nations and internationally. When these indicators are produced on a regular basis, they can show trends over time. They can also reveal changes that follow specific policy actions or school improvement efforts. These changes, of course, cannot be linked with certainty to their causes. However, they do permit analysts to raise hypotheses for more detailed analysis and study. Indicators that have predictive value, e.g. population shifts or numbers of students planning to become teachers, can serve as "early warning signals" of possible conditions in the future. Moreover, to the extent that we understand the links among individual indicators, an indicator system can help explain the conditions and changes they describe.

In essence, information from indicators can inform educational policy and improvement efforts. Indicators cannot provide a precise interpretation of past events, offer clear judgements about present conditions, or point to specific policy remedies for problems that are identified. However, they can generate important new understandings about how education systems are functioning. These understandings can influence how policymakers, educators, and the public think about schooling, and provide useful clues about promising directions for educational improvement. Indicators can serve as tools to aid the ongoing dialogue about what nations want from their education systems and how the goals might best be accomplished. As such, their contribution to educational policy and planning is likely to be substantial.

Most of the available statistics about OECD countries relate to the general characteristics of education systems – to "input" indicators. The two conferences (Washington and Poitiers) made clear the concern among Member countries for the development of a more comprehensive framework for the evaluation of formal educational programmes.

Those attending the two conferences discussed more than 50 possible national indicators and gave examples of the methodological and policy issues involved. These indicators included measures of central tendency and measures of disparity, both providing a basis for comparisons within and among countries.

These indicators can be clustered in a variety of ways, but for the purposes of this proposal, they can be grouped in four clusters:

- input indicators;
- output indicators;
- process indicators; and
- fiscal and human resource indicators.

The Poitiers Conference drew attention to those indicators that could contribute to the development of international comparisons concerning the levels of disparity within different education systems. An appropriate description of the education system needs to take into account the degree of disparity (for example, the gap between the performance of the highest achieving and lowest achieving schools, and between enrolment rates in urban and rural areas) and not just its average characteristics. This is the case for all systems but applies particularly to education where the aim generally is to provide equal opportunity for all. Therefore, it appeared necessary to include disparity measures in the four clusters mentioned above.

The arrangements adopted were appropriate for a first phase of an experimental project as they made use of existing expertise and interest while covering a range of fields. Some fields, notably enrolment and resources, could draw on considerable methodological and conceptual groundwork. On the whole, however, the achievement of the exploratory phase testifies to the difficulties encountered when trying to generate international indicators that are valid and comparable for all Member countries. In the participation and enrolment area, for example, despite the limited conceptual refinements and the need to use existing raw data for calculating indicators, it appeared that the heterogeneity of definitions currently used at the international level is too large to ensure an acceptable degree of comparability. The definition of secondary school completion, for instance, is still ambiguous because there are differences among OECD countries in the way in which secondary schooling is structured, particularly upper secondary education.

Similar problems have been found with respect to the concepts of repetition, attendance, entrance to higher education, completion of higher education and so on. The definitions are those of the International Standard Classification of Education, commonly referred to as ISCED, elaborated by the UNESCO member States in 1978, and several years later by the OECD, in order to harmonize the data published. The crucial aspect of the Classification is not very well settled. The basic terms of reference appear obsolete, vague or missing. Substantial changes are necessary before constructing an international indicator system, pertaining to the validity, consistency and coherence of raw data. In this perspective, revision of the collection instruments and an inventory of problems concerning international comparability of data are urgent.

Great difficulties are also evident in the more speculative and less conceptually developed fields of processes, attitudes, and expectations. Information about teaching staff and expenditures cannot be provided in many countries to permit international comparisons. Similarly, any attempt to combine data on pupils with data on teachers is currently problematic.

The outcomes domain is even more problematic. There is a large body of work on international comparison of student achievement and a number of competing strategies and priorities. The scale of this work is accentuated by growing interest in assessing non-cognitive outcomes and in appraising school system performance through the linkage between education and participation in the economy. These areas also bring with them large bodies of existing work and a variety of conceptual and methodological issues. Maintaining a critical connection between these aspects of the outcome domain is an important policy consideration, but it does not override the need for more focused work on defining and assessing the applicability and acceptability of outcome indicators.

Gaps remain in each indicator cluster. Major work is still needed to conceptualise and refine school process measures, to obtain desirable cognitive and non-cognitive outcomes and to conceptualise and develop the new area of labour force participation. Measures of enrolment, participation, and fiscal costs appear mainly to require technical refinement and operationalisation. For all indicators, development is needed in data collection; data base development, maintenance, and management; displaying and using the results; and making data easily available to countries. A strong conceptual framework to safeguard against the misuse of indicators and ensure a coherent approach to indicator development is also required.

Indicator definition and interpretation still pose substantial challenges. There is marked difficulty in moving from broad statements of agreed concerns and values to

operational definitions of specific indicators. Further conceptual work is also needed to continue to move towards a more complex model of the education system, to strengthen the explanatory power and policy relevance of indicators, and to minimise the data demands of the indicator system by articulating criteria to guide indicator selection and interpretation. Some work should also be done on the periodicity of collection, calculation of indicators and on cataloguing the data strategies which can support indicator development and analysis.

IV. CONCLUSION

Researchers and educational policy-makers should be both open-minded and sceptical where education indicators are concerned. The momentum is nonetheless strong, and the present appears as a period of accomplishment. In a short time, a preliminary set of indicators has been explored that, with planned revisions, holds promise for assessing educational efforts and results. It seems within our grasp to measure and monitor several important means and ends of education. As a result of such efforts, educators and policy-makers can learn more from past successes and failures in using scarce resources – especially human time – to meet complex goals.

Several important effects of the present work can already be identified in the field of educational research and policy: the development and clarification of some of the basic concepts of the theory of education and the stimulation of research; an increased international solidarity, as expressed by the will to co-operate; and the sharing of results and expertise among countries and among the main international organisations active in this domain, including UNESCO, the European Community, and the Council of Europe.

These positive aspects encourage realism in future collaboration. The difficulties and problems involved in such developmental tasks, as well as in the interpretation of the results, have been assessed, and an appropriate operational structure for further work has been established, providing organisational solutions through the use of a diversified but co-ordinated approach.

Indicator systems should not be constructed for their own sake. Educational enterprises are increasingly pressed for effectiveness and productivity, but educational costs and goals that remain beyond immediate measurement are also important. The fact that even some of the more accessible dimensions of education are not being measured at present testifies to the need for a deeper understanding of investments in education.

Modern nations can best review the consequences of recent and future reforms if they monitor and analyse their effects. Educational policy-makers will learn more if such changes are assessed over relatively long time periods and compared with changes in other nations. The permanent monitoring of education systems is increasingly perceived as a means through which democratic societies will be able to maintain, reinforce and equalise the distribution of benefits and responsibilities. Contributing to the achievement of these tasks is one of the key functions and justifications for indicators.

Thus, the greatest attention should be paid to the conclusions that can be drawn from a set of international education indicators. This set should be instrumental in helping to identify the means of improving educational performance. Reforming the curriculum and shaping a general culture in accord with changing socio-economic conditions are among the fundamental policy questions that demand the refined information that education indicators can provide.

Chapter 1

THE FUNCTIONS AND LIMITATIONS OF INTERNATIONAL EDUCATION INDICATORS

by

Desmond Nuttall[1]
London School of Economics
United Kingdom

I. WHY INDICATORS NOW?

Information about the functioning of education systems has been collected for years as a vital tool for planning and monitoring, but the last decade has seen an unprecedented growth in the demand for information and a broadening in its use. Today the concern is as much with quality and access as it is with provision, arising from new attitudes which have influenced and changed educational expectations. The key policy questions have turned from "how many?" and "how much?" to "how good?". As a consequence there has been a search for more sophisticated measures and assessments of educational processes and products, both quantitative and qualitative, and a growing emphasis on fair and valid comparisons within and between education systems for the benefit of policy-makers, managers and consumers alike.

Various factors have led to this demand for more and better information. The concern for improving the quality of education systems stems, in part, from a greater emphasis on value for money in an era of public expenditure constraint. In many nations, educational improvement is being sought by granting institutions greater autonomy, which brings with it a responsibility for accounting for performance, usually in the most objective manner possible. Internationally, economic competition has increased and nations look to their education and training systems to enhance their competitiveness, and thus to comparisons between systems to identify possible improvements.

This move towards more rational and open policy formulation needs to be underpinned by a clearer understanding of how education systems are organised and function, so that information and judgements about them can be more helpful. In return, this requires the development of new evaluation methods which might in themselves influence the frame of reference for the educational debate.

It is in this context that development work on indicators is particularly needed. First, such work needs to address the place of indicators in evaluation theory, and how they can complement and supplement judgements by inspectors or independent evaluators, possibly through integration into expert systems. Second, such work needs to overcome the

limitations of earlier attempts to develop indicators, which resulted in a lack of utility for policy-makers and managers.

These limitations arose principally from inadequacies, especially lack of comprehensiveness, in the underlying models of education: some concentrated only on the economic dimension of education while others limited themselves to the financial and human resources invested in education or just to outcomes, without regard to educational processes or the local or national context. Some were even more limited, using information that happened to be readily available in an *ad hoc* fashion. These limitations are documented in Chapter 2 by Max van Herpen[2].

In the light of this experience, it is clear that new work on education indicators must be conceptually more rigorous, using information which is soundly based, related to a valid representation of the education system and must produce measures which are comprehensive and comprehensible. The limitations of indicators must be identified and acknowledged.

The aim of the following section is to amplify the broad conceptual framework and rationale for the CERI/OECD Project and to offer a set of principles, both technical and political, to guide its work.

II. DEFINITION AND NATURE OF INDICATORS

What is an education indicator?

An education indicator tells something about the performance or behaviour of an education system and can be used to inform educational decision-making. Not all education statistics qualify as indicators. Like economic and health indicators (e.g. GNP, infant mortality rates), education indicators focus on key aspects of their topic. Consequently, indicators are expected to reflect the condition of the system as a whole, or of some significant part or element of the system.

To be an indicator, an education statistic must also have a reference point against which it can be judged. Usually the reference point is some socially agreed-upon standard (e.g. a minimum reading age to indicate basic literacy), a past value (e.g. the 1970 level of mathematical attainment), or a comparison across schools, regions or nations. Obviously, indicators do not tell everything about education systems. Instead, like economic or health indicators, they provide an "at a glance" profile of current conditions.

Characteristics of indicators

Indicators are designed to provide insight into the behaviour and effectiveness of education systems and to be useful in a policy context. To meet these goals, indicators must satisfy a number of substantive and technical criteria. For example, indicators should provide at least one of the following types of information:

- Information that describes the system's performance in achieving desired educational conditions or results. This information helps to describe the system's current functioning and effectiveness.

- Information about features of the system known to be linked with desired conditions and outcomes. This information can help policy-makers, educators, and the public predict future performance.
- Information that describes enduring features of the system. This information helps policy-makers and educators better understand how the system works and to assess the implication of changes over time.
- Information that is relevant to educational policy. This information provides insight into current or potential problems that are of particular concern to policy-makers or that are susceptible to change through action.

These criteria emerge from a review of earlier efforts to develop social indicators. That experience demonstrated that technical quality is not sufficient to guarantee their use and continuation. Indicator systems must also produce information useful to the policy community if they are to survive as publicly-supported endeavours. Appropriate criteria for the CERI/OECD Project are put forward in Section IV of this chapter.

Indicator systems

Given the complexity and diversity of education systems, it is obvious that an individual indicator conveys limited information. To compensate for the unidimensional nature of each indicator, it is necessary to build a system of indicators, that is, a coherent set of indicators that together provide a valid representation of the condition of a particular education system, not just an *ad hoc* collection of readily available statistics. Ideally, an indicator system will provide information about how individual indicator components work together to produce an overall effect. In other words, the policy and interpretative value of all the information to be gained from a system of indicators is greater than the sum of its parts.

To provide this overall picture, the selected indicators should be logically or empirically linked. The linkages should proceed from a model or framework that describes how the education system works. The model by itself permits the broader assessment of an indicator's relevance. Drawing on the model, combinations of indicators can be used to diagnose current and future conditions. Further, particular combinations of indicators can provide clues about what is wrong and what changes might lead to improvement.

In the development of a system of international education indicators, the choice of a suitable model is particularly difficult since the educational and political values of different nations vary, as do their institutions and procedures. But there is a fundamental similarity, at least among the systems of OECD Member states, and the framework for the model adopted by the Project is of general applicability (see Section IV). Nevertheless, there are sufficient differences between systems to require careful interpretation. Each nation is influenced by a different set of contextual factors and it is important to account for these differences when making comparisons.

III. FUNCTIONS AND PURPOSES

The broad purposes of indicator systems are, in the words of Oakes (1986), "to measure the health and effectiveness of the education system and help policy-makers make better decisions". The allusion to a medical check-up is appropriate since it draws

attention to the skill of the practitioner in interpreting and making sense of the information available from many sources (such as blood pressure, pulse rate and body temperature) – information which can help in diagnosis but is not prescriptive or judgemental.

These broad purposes can be viewed from two perspectives, that of the researcher and that of the policy-maker (applying this term to people at every level of the system from a national ministry to an individual institution). The policy-maker is interested in monitoring and evaluating the quality of education, and developing policy options in the light of the information available. The researcher is interested in the description and comparison of education systems, and hence in improving the analysis and understanding of education systems – in short, in increasing knowledge and understanding (which may also be of value to the decision-maker).

Furthermore, these purposes apply at different levels within education. Information may be used to evaluate or understand the performance of different institutions, or different regions or groups (such as males and females) within a national system. Data at the national level are widely used in educational debate and are commonly collected and analysed by governmental agencies. The final level involves direct comparison of national indicators between countries, and stimulates debate about possible causes of the differences against the background of national contexts and differences.

The following paragraphs look in more detail at these purposes of indicator systems and provide some examples of indicator systems in use. Within the Project, all levels are considered since the aim is not only to develop an international indicator system, but also to foster the exchange of information about national and local indicator systems and their application and use.

Uses for the policy-maker

Monitoring and evaluating

There are many models of educational evaluation, but their common feature is a requirement for information which may be quantitative or qualitative, reserved for administrative use or widespread within the public, monolithic or pluralist. In no case is the information more than an aid to judgement, and the modes of analysis and presentation can significantly enhance or detract from its value.

Within countries, the use of indicators in the evaluation process is widespread through such mechanisms as national or state testing programmes, consumer attitude surveys and surveys by school inspectors, as well as through the routine collection of statistics on enrolments, teacher supply and so forth. Recent years have seen a rise in demand for such information, both by the policy-maker and the consumer, and an increased sophistication in its analysis, looking at disparities and dispersions within systems (not least to study the performance of disadvantaged groups) as well as the general level of performance. This interest has been mirrored internationally with a growth in the number of countries participating in IEA[3] surveys of educational achievement, in many cases sponsored by governments, and the rapid growth of the surveys by the International Assessment of Educational Achievement.

Although many indicators provide valuable descriptive information about education systems, for example to discharge public accountability, most users of indicator systems are usually more interested in comparisons and judgements of quality, to answer ques-

tions of "how good?" rather than just "how much?". It is fears about crude comparison, involving the creation of league tables, that generate the most anxiety whenever indicator systems are introduced.

Comparisons can, however, be carried out in many different ways. The most familiar comparison in education is between pupils through norm-referenced scoring or grading systems. The development of student assessment and evaluation in the last 20 years has stimulated alternative perspectives: comparison against a defined standard, as in many proficiency tests such as the driving test, or comparison against one's own previous performance. These ideas are equally valid in the case of institutional or systemic indicators, and their use can avoid unfair comparisons between unlike systems.

An example of the potential value and drawbacks of comparisons in stimulating improvement comes from the United States airline industry; each airline now has to report how many of its flights are delayed for more than 15 minutes. The number of delayed flights has fallen substantially, to the benefit of the passengers. The pressure generated by the publication of such indicators is not without danger. Of course, passenger safety should not be jeopardised, nor should passengers on delayed connecting flights receive less consideration. So other indicators – parts of an indicator system – are also essential if the comparisons are to be fair and comprehensive.

Developing policy options

Monitoring and evaluation may in themselves have value, not least in encouraging those responsible to report upon their stewardship. However, their real use is in stimulating improvement through the light that they shed upon alternative strategies. Comparison of policies and procedures in different schools, states and nations is, in this era of instantaneous global communication, widely adopted by politicians and their advisers. Its effectiveness is enhanced if it is carried out in the light of information that reflects the relative success of the policy or procedure against internal and external criteria, including appropriate international comparison. This is commonplace in the field of economics where international comparison of information from indicators such as inflation and money supply are used to generate policy options. The OECD's very existence is testimony to the value of the international exchange of ideas, and there are examples from other facets of its work which point to the value of comparative information as a stimulus to policy analysis and policy development. Indicator systems for scientific and technological research and development provide an excellent example.

Uses for the researcher

A common research approach is to model education systems through the use of indicators, to develop a deeper understanding of a phenomenon. One good example is the study of school effectiveness[4]. Many studies throughout the world have analysed the factors that appear to contribute to the relative success of some schools, usually measured in academic terms, but often more broadly based, in comparison with others apparently very similar in their student body and other characteristics (such as size) outside their control. Current research in the field is leading to a more sophisticated understanding of effectiveness, and the extent to which some schools may be equally effective for all students irrespective of race, sex or class, while others, ostensibly effective overall, may be less effective for some groups than for others.

The use of indicators in comparative studies is also common in research, but is viewed by most as simply a starting point for analysis. Interpretation of the indicator figures can be done only in the context of the values, objectives and structures of the particular education system, and hence best by local actors imbued with a full understanding of that system.

Conclusions

These examples of indicator systems in use for a variety of purposes support the claim that indicators have much to offer the education system in promoting understanding and in devising and evaluating policy options. They also emphasize the fact that indicators are open to misinterpretation and potentially to misuse.

This is particularly true when too much is expected of indicators – when they are uncritically viewed as objective yardsticks, even arbiters, rather than as partial and not perfectly reliable measures which can be no more than suggestive, diagnostic and capable of improvement and development.

These limitations must be openly acknowledged, especially to underscore the fact that indicators only support judgements and are not substitutes for decision-making:

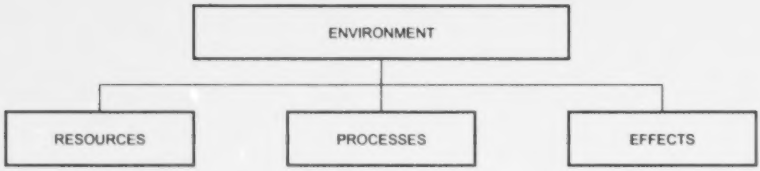
> "In the end, interpretations of what indicators mean and decisions about what policies should be implemented will be influenced by values as well as by knowledge. Indicators cannot remove fundamental decisions about schooling from the political process – from the reflection and debate among policy-makers and the public ultimately responsible for its healthy functioning" (Oakes, 1986, p. 23).

Nowhere is this warning about the limitations of indicators more apt than in the field of international indicators. The interpretation of indicators must be done in context, against the background of national goals and values – the figures cannot be permitted to "speak for themselves".

In addition, caution in interpreting indicators is advisable as a result of possible lack of a valid match between the indicator system and the implicit or explicit models of the education system underlying it. Not all indicators will have as well-established a link with the underlying concepts as the length of mercury in a thermometer has with temperature. More than that, the inter-relationship of concepts in the model itself may also be more speculative than the relationship of body temperature to other concepts in differential medical diagnosis.

IV. THE CERI/OECD PROJECT

The aims of the CERI/OECD Project are set out in Section I and embrace both international comparisons and international co-operation in the development of better local indicator systems. Given the criticisms of earlier work, a comprehensive framework for the development and interpretation of indicator systems was essential, and the Project adopted the systemic scheme set out below:



Any understanding of the effects or outcomes of education must be informed by the educational processes employed and the resources (fiscal and human) deployed, against the background of contextual factors in the environment of schools or of education systems. Such factors may be economic, demographic, socio-cultural and socio-psychological (for example, public expectations of and satisfaction with the education system). This framework has reasonable claim to be comprehensive, but does not obviate the need for interpretation of indicators in the broader context of different educational goals and priorities. Furthermore, it allows the development of different models of the educational process; the original Working Plan for the Project provided possible examples based in part on the proposals from the Washington and Poitiers conferences.

Under the general direction of the Scientific Advisory Group, the work of the preliminary phase of the Project has been undertaken by five Networks of interested countries focusing on more circumscribed aspects of the general framework, namely:

i) flow of students through the education system;
ii) students' outcomes, including attainments;
iii) schools and their environment;
iv) the costs of education;
v) attitudes and expectations.

It is recognised that these five aspects do not comprehensively cover all domains of the framework and that, because of varying membership of the Networks, the indicators developed in the preliminary phase cannot be assembled into a full international set. Rather, in this phase the aim has been simply to examine the feasibility of collecting and analysing a range of international indicators in preparation for a comprehensive approach.

In light of the analysis of the strengths and limitations of indicators as set out above, certain general principles emerge to guide the work of the Project, to ensure that the strengths are exploited and the dangers minimised. These principles are as follows:

a) Indicators are diagnostic and suggestive of alternative actions, rather than judgemental.
b) The implicit model underlying a set of indicators must be made explicit and acknowledged.
c) The criteria for the selection of indicators must be made clear and related to the underlying model.
d) Individual indicators should be valid, reliable and useful.
e) Comparisons must be done fairly and in a variety of ways (e.g. with like groups, with self over time, and using dispersions and differences between sub-groups as well as averages).
f) The various consumers of information must be educated about its use.

The principle *(a)* guides the whole work of the Project, and will be referred to at every appropriate point in the presentation of results. Similarly, principle *(f)* is all-

pervasive and will guide the drafting of any commentary on the Project, or on particular results.

The importance of validity, and of an explicit relationship between the indicator system and the underlying model of the educational process, has already been stressed. Principles *(b)* and *(c)* are designed to reinforce this vital point, and to aid the interpretation of indicators produced.

The technical quality and utility of individual indicators are upheld by principle *(d)*. More specifically, it is proposed that each indicator should meet the following criteria:

- Provide information that describes central and enduring features of the education system;
- Provide information about current or potential problems;
- Provide policy-relevant information;
- Measure observed behaviour rather than perceptions, wherever possible;
- Generate data from measures that are generally accepted as valid and reliable;
- Be feasible in terms of timeliness, cost, and expertise.

Principle *(e)*, about comparisons, also offers a vital safeguard against the misuse of data. It is suggested that a special study of the most appropriate ways in which to present international education indicators should be carried out, reviewing in particular the experience of IEA and other relevant experience (e.g. the presentation of test results and other indicators state by state in the United States).

V. CONCLUSION

In the first phase of the Project, each Network has investigated the feasibility of deriving and calculating a set of international education indicators, but the progress made by each Network has varied according to the complexity of the areas studied and the amount of existing work that could be built upon. In all Networks, the Project has at the very least provided a valuable forum for the exchange of information about methods and practices of developing and using education indicators for monitoring, policy-making and managing education systems.

The first phase has confirmed that it is not easy to select indicators from any of the domains to provide a set that will be perceived as valid by every participating country. Although there is broad agreement on the goals of education, there are inevitably differences in the values surrounding those goals; moreover, each goal may have a different significance and importance in each country. In many cases, these differences do not prevent the calculation of an indicator but will mean that the interpretation of the international set of indicators inside each country will rightly be somewhat different. As long as the first principle is accepted, namely that indicators are diagnostic and suggestive of alternative actions, rather than judgemental, such differences in emphasis will not cause difficulty.

There have been some occasions, however, where there was no agreement on the creation of an indicator because the values and procedures in different countries were so different that no consensus could be reached. For those countries that attach significance to such a missing indicator, the final set of agreed indicators must lack validity.

It is clear, therefore, that there is a need for further conceptual work to understand better the process of deriving a valid set of indicators from a broad and accepted conceptual model of education. This work should also embrace a study of the most effective operational steps in the process of indicator derivation and validation.

NOTES

1. I would like to acknowledge my debt in the development of this chapter to the members of the Scientific Advisory Group of the Project and the Network Chairs together with the staff of OECD/CERI.
2. Max van Herpen's contribution provides a range of examples of indicator systems and the models which support them. The issue of validity, referred to in Section IV of this chapter, is also explored in more detail.
3. The International Association for the Evaluation of Educational Achievement (IEA) is an international, non-profit scientific association incorporated in Belgium for the principal purposes of:
   i) undertaking educational research on an international scale;
   ii) promoting research aimed at examining educational problems in order to provide facts which can help in the ultimate improvement of education systems;
   iii) providing the means whereby research centres in the various member countries of IEA can undertake co-operative projects.
4. See Chapter 3 by Jaap Scheerens for further information on school effectiveness.

BIBLIOGRAPHY

Fuhrman, S. (ed.) (1988), Series of articles on education indicators, *Phi Delta Kappan,* March, pp. 486-513.

Murnane, R.J. (1987), "Improving education indicators and economic indicators: the same problems?", *Educational Evaluation and Policy Analysis,* Vol. 9, No. 2, pp. 101-116.

Oakes, J. (1986), *Educational Indicators: a guide for policymakers,* Santa Monica, CA, The Rand Corporation, for the Center for Policy Research in Education.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N. with Picus, L. (1987), *Indicator Systems for Monitoring Mathematics and Science Education,* Santa Monica, CA: The Rand Corporation.

Wyatt, T. and A. Ruby (eds.) (1988), *Indicators in Education,* Sydney: Australian Conference of Directors-General of Education.

Chapter 2

CONCEPTUAL MODELS IN USE FOR EDUCATION INDICATORS

by

Max van Herpen
Researcher, Central Statistical Office
the Netherlands

I. INTRODUCTION

The aim of this chapter is specifically to examine the range of conceptual models for education indicators that have been thus far developed, either explicitly or by implication.

The array of models that exist is testimony to the differing perceptions and purposes of both models and indicators. A number of factors can be at work here:

- The unit of analysis may be different. At the micro-level, learning and teaching are looked at as activities; at the meso-level, the school is seen as an organisation; at the macro-level, education is seen as a functional system in society. These different levels are closely bound up with the different disciplines involved in studying the educational process – psychology, sociology and economy – each having its own conceptual frames.
- The scope of the models may be different. Sometimes only a sub-system is looked at, e.g. higher education. Sometimes the education system includes only formal education. An important question is then where to draw the boundaries of the education system.
- Indicators are expected to do a number of different things: they monitor, that is, they keep an eye on the areas considered relevant, or they evaluate concepts like quality, performance or "health". These are quite diverse functions which call for different types of indicators.
- Different indicators are also needed when it comes to evaluating the education system, educational policy, a reform programme, or teaching.
- Models may be simple schemes in which different areas or elements are distinguished or they may have a completely causal character.
- The same concepts, indicators and variables are defined by different terms and even under the same terms different things may be understood. Also the distinction between a concept, a variable and an indicator is not always made in the same way. The term "indicators" should be used in an evaluation context, and the term "variables" in a measurement context. This distinction is not always

made. Sometimes variables are included in the model, sometimes they are not. Consequently the different models are difficult to distinguish and a standard nomenclature is urgently needed.

What then is a conceptual model? A conceptual model specifies the relationships between a set of related concepts. A conceptual model is the first step in formalising a theory. Although there are numerous theories concerning education, e.g. the human capital theory, different learning theories, theories of effective schools, and sociological theories about the social function of education, there are no theories of education systems as such. There is no theory like macro-economics that explains the relations between the different educational concepts on a macro-level. Should all these different theories be incorporated into one global theory for the sake of education indicators? Indicators such as test scores relate to "learning theories" and others such as "rate of return" belong to the human capital theory, while still other indicators seem to belong to the "effective school" theory. Or would it be necessary to develop a different theory for every indicator? Or is it sufficient to ground each indicator in an appropriate body of theory?

In his review of the educational part of "Social Indicators 1979", Mare (1981) states that no single conceptual scheme is appropriate for education indicators, because the education system is multidimensional.

> "Of the many dimensions that can be enumerated, three seem central. First, the formal school system is the main locus for the teaching of cognitive skills and practical and theoretical knowledge and for the transmission of values, attitudes and behaviours. Second, schools are 'people-processing institutions' in which the population passes through a sequence of well-ordered, age-graded levels. Thus the educational system is an agent in the demographic process by which socioeconomic inequalities are perpetuated from generation to generation. Third, the educational system is a complex organization. Supported largely by public resources, characterised by enormous variation in organisational forms, employing a sizeable fraction of the labor force, and enjoying unprecedented growth in the twentieth century, schools have a political and economic life of their own".

The three dimensions are thus: demography, performance and organisation.

By contrast, Averch *et al.* (1971) in their review of the research on effective schools see five different approaches used in educational research.

i) The *input-output approach* assumes that students' educational outcomes are determined by the quantity and quality of the educational resources they receive. This approach views the school as a black box presenting the empirical problem of how to establish the relation between input and output. In practice, cross-sectional data are used, and the effects of input variables are estimated by multiple regression.

ii) The *process approach* attempts to examine the processes and methods by which resources are applied to students. Here the researcher looks inside the box: resources are assumed to be fixed and what matters are the processes by which the resources are applied to the students and the students' response to the processes.

iii) The *organisational approach* supposes that better educational outcomes for individuals result from improved functioning of the organisations that deliver education. The school is seen as having to adapt to the needs of the changing set

of students and to a changing set of outside pressures. Consequently, focus on the output side is on determinants of innovativeness and responsiveness, and focus on the input side is on the rules, incentives, procedures, leverage, and so forth. The measure of responsiveness to change is the ability to adapt to a changing clientele. The assumption is that responsive schools will deliver satisfactory outcomes: this does not necessarily mean the highest possible level of academic standards, since schools have multiple objectives which are not limited to academic outcomes. Thus, schools will be perceived as doing well if their academic achievements are satisfactory and their other goals are also satisfied. The perceived crisis of the classroom is caused by inflexibility in the face of changing demands by students, parents, the community, and the government. The purpose, then, is to understand the behaviour of the whole system: to describe the shape of the box and how and what happens to the people in it – not just the students, but the teachers, the administrators, and the education community as a whole.

iv) The *evaluation approach* attempts to analyse the effectiveness of broad educational interventions that are directly related to large social policy issues. Its primary concern is to identify the relationship between the existence (or magnitude) of an intervention and educational outcomes.

v) The *experiential approach* is concerned with what happens to students in schools. The school is viewed as an institution containing students and having an impact on them. Although the school's impact may affect educational outcomes, this is not viewed as the primary concern – considerable stress is placed on the impact as an end and outcome in itself. The primary emphasis is on the effects of school experiences on students' self-concepts and on their relation to other people and to social institutions.

Is it possible and useful to integrate these different approaches into one comprehensive model or will one or more different approaches have to be chosen at the expense of the others?

What follows is a comparative overview of different models advanced in the readily available literature to aid and inform the choice of a conceptual model in an education indicators project. The division of input, output, process and context is common but models vary in their focus and scope. Since a large number of models is presented, no detailed descriptions will be given. These can be found in the original source material. The focus is on the macro-level, although meso- and micro-levels will be briefly discussed. A significant constraint on the critical quality of the review is the relative lack of information on the actual use of the different models and the users' evaluation of appropriateness.

The following description of models is drawn from four kinds of reports:

1. Concrete indicator reports, of which there are only three:
   - *The Condition of Education* (National Center for Education Statistics, United States, 1978).
   - *Educational Indicators for the Elementary and Secondary Levels* (Ministry of Education, Quebec, 1988); and
   - *Educational Indicators in Japan* (Yamamoto, 1987).
2. Reports of systems that seem to be in an advanced state of realisation: Council of Chief State School Officers, France, Netherlands, California.

3. Reports from international organisations making elaborate proposals which up to now have not been realised: Johnstone (1981), OECD (1973*a*), United Nations (1975).
4. Reports with more or less the character of a research report, in which a concrete indicator system was presented that had not been followed up: Olkinuora (1973); Conrad (1973); Delors (1971).

In analysing the different conceptual schemes, a number of clusters are distinguished and presented in the following paragraphs under the headings:

- Social indicators of education;
- The inductive approach;
- Policy goals;
- Manipulable input variables:
- Division between consumption and production;
- Compound indicators;
- The education system as a whole;
- The school as the unit of analysis.

Annexes 1 and 2 give a systematic and chronological overview of the different models.

II. SOCIAL INDICATORS OF EDUCATION

Following the success of economic indicators, there was an enormous interest in social indicators, resulting in the social indicator movement [see e.g. Duncan (1969) and Land (1975)], and some of the social indicators were devoted to education. A certain autonomy gradually developed in the different fields, giving rise to indicators for systems such as science, health, environment and education.

A distinction should be made between education indicators and social indicators of education. Education indicators are aimed at the performance of the education system, while social indicators of education try to measure an aspect of educational well-being. Of course there is a relation between the two; social indicators of education should be a part of education indicators.

The OECD (1982) distinguishes two areas of indicators. The first area is use of educational facilities including such concepts as: formal education (average number of years of formal education experienced by the population); adult education (the percentage of the population who have participated in adult education during the previous year). The former indicates the level of education attained in formal education, the latter the participation in education by adults at the present time. Two related concepts in this area are physical and financial accessibility. A second area is called learning, which draws attention to what is learned in education, including indicators such as the literacy rate. A third area which is mentioned, but not further developed, is "satisfaction with educational experience".

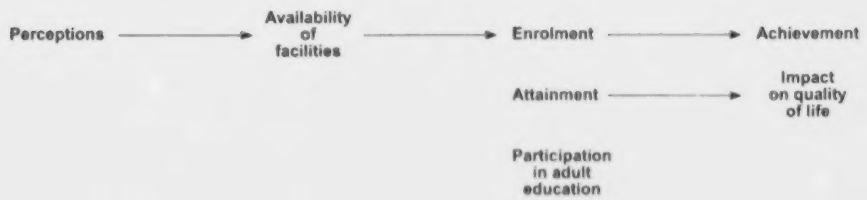
The different concepts could be modelled as follows:

The US Bureau of the Census *Social Indicators III Report* (1980) also makes the distinction between social indicators of education and education indicators and includes both kinds of indicators. The indicators for education are presented in five sections:

1. *Public perceptions:* the public's commitment to education; e.g. perceived importance of college education;
2. *Resource commitments:* e.g. pupil/teacher ratios, public and private;
3. *Enrolment and attainment;*
4. *Performance and achievement;*
5. *Adult education and training.*

"Attainment" in the Report's Section 3 is seen as an indicator of the system's output. In order to assess the impact of education on the quality of life, it is essential to examine outcomes. These can be measured at least at two levels. The first level is what the students learn while in the education system. The second level is related to their performance and achievements in the wider social system. Indicators at the first level are the percentage below the modal grade (repeat rates) and achievement tests. Indicators at the second level include income by educational attainment and under-employment of bachelor degree holders.

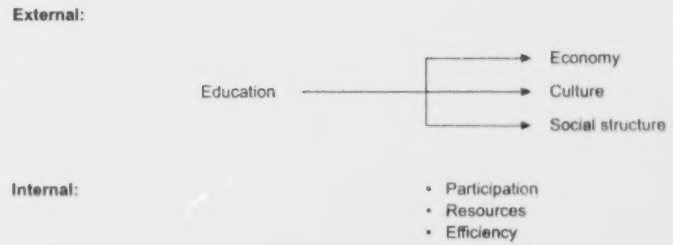
The conceptual frame in this report is an extension and a modification of the model used by the OECD (1982). "Use of educational facilities" is now called "enrolment and attainment" and "Adult education and training". The distinction made between output (i.e. attained educational level of the population) and outcome (i.e. educational achievement) and impact of education on well-being (i.e. income, work, housing, health, participation in community and cultural activities) is interesting. The conceptual frame of *Social Indicators III* can be modelled as follows:



In an earlier attempt, Delors (1971) tried to develop a system of social indicators around 21 themes including education. Education has a function referring to the past: transfer of knowledge, skills, etc., and to the future with education as the basis for future development, new knowledge, attitudes. Three sub-sectors are predominantly important for education: the economy, the social system (the social stratification) and culture (attitudes). Delors distinguishes two kinds of indicators:

1. *Performance indicators* which relate to the main functions of education for the economy, the social system and culture;
2. *Indicators for the internal functioning of the education system,* covering participation, resources, and efficiency measures such as: pupil/teacher ratio, part of educational budget spent on educational research, teacher training and information, and recurrent expenditure per pupil.

Delors uses two rudimentary models to describe the relationships between the indicators:



A more structured model based on flow matrices and mathematical modelling was advanced by Stone in his report for the United Nations: *Towards a System of Social and Demographic Statistics* (United Nations, 1975). For each part of the system described in the report, a set of indicators is proposed next to the statistical system. The education indicators are divided into three groups:

- stocks and flows of students;
- inputs and outputs of educational services;
- the educational attainment of the population.

III. THE INDUCTIVE APPROACH

This section presents a number of concrete indicator reports that seem to have a practical orientation. Instead of a theoretical framework, the starting point is the availability of statistics; hence, the approach is mainly inductive. The difference between statistics and indicators here is rather small and the concepts are at a low level of abstraction.

A recent example of a concrete indicator system is the annual publication of *Educational Indicators* in Quebec. Three major aspects are evaluated: the funding of education, the availability of educational services, and the relevance of students' learning achievements. The booklet comprises five sections and in 1987 had eighteen indicators, in 1988 twenty-one. The five sections are the following:

1. *Spending;*
2. *School careers* (probability of entering certain stages, repeat rates and drop out);
3. *Examination results;*
4. *Graduates* (persistence until graduation, type of study, transition to higher education and work);

5. *Adult education* (spending and graduates).

The sections follow the sequence of input-throughput-output of the formal primary and secondary education system and ends up with the non-formal system.

On a regular basis since 1969 Japan publishes an *International Comparison of Educational Indicators* to compare Japanese conditions and levels of education with other countries. Three different sections are covered: Diffusion of education (enrolment and advancement rates), Teachers (pupil/teacher ratios and the ratio of female to male teachers), Educational finance (ratio of educational expenditure by government to GNP, to the total of government expenditure, expenditure on school education by government and by parents, details of expenditure on education, tuition fees of universities and colleges, scholarship). The model is a three-way partition of pupils, teachers and finance, a partition common in education statistics. A similar division is made by a Technical Committee on Education Statistics, Canada (1988) in which the three elements are divided into an input, throughput and an output indicator. For each cell in the scheme below an indicator can, in principle, be produced. Concrete examples are given in the report.

	INPUT	THROUGHPUT	OUTPUT
Pupils			
Teachers			
Finance			

A slightly different approach is made in a UNESCO paper (1976*b*) in which an inventory is made of education indicators in use in different countries. For least-developed, developing and developed countries, a separate fourfold frame is designed.

1. Educational attainment of the population (median years of schooling, proportion completing);
2. Use and distribution of educational services (enrolment);
3. Educational achievement;
4. Inputs, outputs and performance of educational services (pupil/teacher ratio; median level completed by teachers; total consumption expenditure per pupil, percentage of GDP; gross fixed capital formation in education as a percentage of total gross fixed capital formation).

IV. POLICY GOALS

In the models that are presented in this section the accent is on output of the education system as defined by the political system.

One of the earliest reports in this area is the OECD's *Indicators of Performance of Educational Systems* (1973*b*). This report tries to develop indicators by defining goal areas. A goal area is defined as an area in which society has continuing interest or concerns, and to which education is related. It suggests five major goal areas:

1. *Transmission of knowledge and skills:* Individuals should be able to function more or less autonomously with respect to all the major institutions.
2. *Education and the economy.*
3. *Equality of educational opportunity:*
   a) Provision for the disadvantaged;
   b) Education and the distribution of income.
4. *Provision of educational services for individual requirements.*
5. *Education and the quality of life:* Education must contribute to the non-monetary aspects of well-being. Different aspects are considered: health, participation in work and leisure, and the extent to which education contributes to the realisation of human potential.

A subsequent UNESCO report (1976*a*) presents more or less the same model. Arguments in favour of the model include:

- The impact of demographic change on education systems is of importance.
- The interrelation between the education sub-system and the sub-systems of work and health is also of importance.
- A clear distinction between input, process and output is necessary.
- Diagnostic goal-directed indicators are important, as is system information, so-called variables, if they are incorporated in a model.
- As no clear set of educational goals is available from which indicators can be deduced, the different outlooks of the psychologist (the learning process), of the economist (education as an investment), the sociologist (theories about access to education and the effect on social class stratification), of manpower forecasts and attitude surveys (estimates of private demand for education, questions on the quality of life) should be combined.

Olkinuora (1973) also advocated a mixed model combining goal areas and "the manipulable basic entities in educational planning". His defined goal areas are:

- educational equality (social stratification);
- relevance from the point of view of production needs (economy and labour market);
- relevance from the point of view of the individual needs and resources (quality of life);
- educational efficiency and costs (accounting, cost/benefit).

These goal areas resemble those of the OECD (1973*b*).

The manipulable basic entities of educational planning are divided into quantitative (structure of services, flow of students and education stocks) and qualitative (nature and content of education and educational achievements).

For each of the 20 cross-sections, indicators have to be developed. Olkinuora gives concrete examples for each of these based on the Finnish experience.

	EQUALITY	ECONOMY	WELL-BEING	EFFICIENCY
Structure				
Flows				
Stocks				
Content				
Achievement				

After the Second World War, there was a general feeling in Germany that the education system had failed because it was authoritarian and out-of-date, and had contributed to the rise to power of the Nazi regime. One of the first things to be done in Germany was to build a new education system. Modernisation and democratisation were the two main goals of education policy. Conrad (1973) tried to find quantitative measures for these two important qualitative aspects of the school system: democratisation and modernisation, and to it he added a financial indicator – readiness for expenditure. Democratisation is looked at in three aspects: do different ages, different socio-economic groups, and both sexes have equal access to education? Modernisation is differentiated in three aspects: internal and external efficiency, and intensity of education. Internal efficiency indicators relate to drop-out, graduation and transition rates; external efficiency indicators relate to graduation rates; intensity indicators relate to class size and pupil/teacher ratios. Well-known monetary indicators are used, such as expenditure per capita, per pupil, the proportion of the budget and GNP, investment and expenditure for personnel.

An indicator system related directly to political goals has been recently introduced by the Netherlands' Central Bureau of Statistics (CBS) in response to the decision to implement a more comprehensive type of school. The five areas for which a number of indicators are to be developed are:

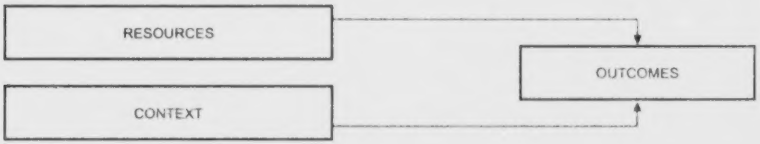
1. Level of educational attainment (this must be raised);
2. Equality of opportunity (must be more equal);
3. Common cultural basis (must be strenghtened; the system is now highly selective, so that children of different aptitude and different background have quite different educational programmes);
4. Educational costs (the new situation must not cost more than the old one); and
5. Educational supply (it must be accessible to a greater number of students and one consequence of the reform will be an enlargement of scale).

Each area is divided into a number of aspects for which indicators are developed.

V. MANIPULABLE INPUT VARIABLES

The United States' contribution to the Washington Conference (OECD, 1987) states, "The selection and combination of a number of indicators into a system for reporting to the nation is influenced by a model of how education works". A basic principle guiding indicator development in the United States has been the need to measure outcomes of education and those inputs, processes or context that have an influence on outcomes. Context is that which you have to take as given, it is not manipulable – input and process are. That is why models can differ. At the state level, policy is an input but at the federal level it is a contextual variable. A new element is introduced here – the distinction is made between manipulable and non-manipulable factors that have an influence on outcomes. This view is taken with a policy-maker's perspective. The policy-maker can be found at different levels in the system: at the state level, the regional level and the school level.

For the international comparison, the United States proposes a three-stage model.



Resources refers to elements such as pupil/teacher ratios. Context includes diverse factors such as total time spent on school work and student drug abuse, while Outcomes-looks typically at achievement and participation.

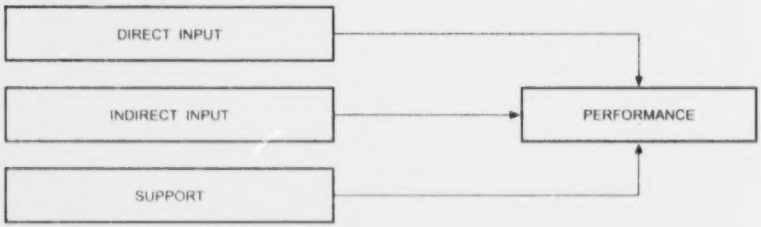
This model is also used by the National Center for Education Statistics (NCES): *Indicators of Education, Status and Trends* (1985) and later in *The Condition of Education* (1987*a*), and by the "Wall Chart", a table released by the Secretary of Education of the United States in 1984. The publication of this "Wall Chart" seems to have been one of the reasons for the recent interest in education indicators. For a critical appraisal of this "Wall Chart", see Wainer *et al.* (1985).

The distinction between manipulable and non-manipulable variables is also made by the Urban Institute in its project "Selected Education Indicators for Twenty-one Major Cities: Some Statistical Benchmarks" (Baker, 1978). This was a preliminary project which experimented with descriptive indicators. The objectives of the study were: *a)* to develop and present statistical benchmarks which could aid in the evaluation of education in the larger cities; *b)* to illustrate and discuss how these measurement techniques could be utilised to aid in such evaluation; *c)* to suggest where further research and data-gathering efforts were needed. The conceptual framework used involved a fourfold education indicator classification:

1. *Educational Performance Indicators:* The educational results that are achieved in the community.
2. *Direct Educational Input-Related Indicators:* Those factors which are believed to contribute to the education of youth and over whose allocation the education authorities have some control.

3. *Indirect Educational Input-Related Indicators:* Those factors which are believed to contribute to the education of youth, but which are not usually the direct responsibility of education authorities.
4. *Educational Interest and Support Indicators:* The degree of commitment which a society makes to the education of its youth; the proportion of resources it devotes to education.

This model could be depicted as:



Only available data were used. There was surprisingly little correlation between the input and output indicators and no causal relations could be assumed. One weakness was the lack of more diverse, valid and reliable educational output indicators. "Better output measures would not tell us how to correct deteriorating conditions. However, we could at least keep track of improving and deteriorating areas".

The United States Council of Chief State School Officers (CCSSO) also concentrated on manipulable variables and expressed the desire for improvement in developing an indicator system. The CCSSO framework for education indicators requires the design of a common set of data through which valid and reliable comparisons can be made (between states and the nation; among states). The focus is at the state level. Three components are discerned in the model:

1. *Contextual factors:* covering notably demography and level of resources;
2. *Policies and practices:* referring to instructional programmes, resource allocation, and policies on participation in the state; and
3. *Educational outcomes:* dealing with attendance, school completion, student achievement, and status after secondary education.

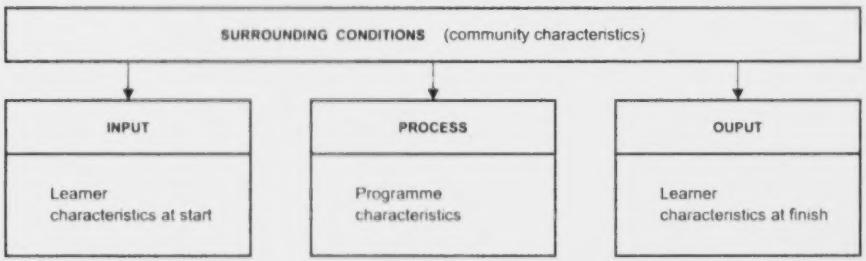
Outcomes are defined as the results of the education system. These outcomes occur as the consequences, in part, of contextual factors which cannot be changed, such as the demographic character of a given state or locality, or the characteristics of its student population. Finally, the outcomes are seen as resulting from educational policies and practices which can be changed. This framework allows accomplishments or achievements to be monitored as the results of decisions made by policy-makers, but it also allows for outcomes to be considered in terms of the conditions under which school systems operate.

In the presentation of the indicators, contextual factors should be used to group or cluster states. In this way states confronting similar circumstances can be compared. A classification system should be developed so that states can be classified by contextual factors. A similar approach is used by the state of Oregon (Saalfeld, 1973) where an

input-output model was applied, and the state compared with other comparable states. This model could be depicted as:

with context variable assumed to be constant or of common effect.

A last example of an education indicator system in which there is strong interest in manipulable variables is the project of the Bureau of School Program Evaluation (BSPE) (see Baker, 1978). BSPE developed a model for performance indicators for school districts in the state of New York. Two types of input are important for student achievement: those which are insensitive to policy control, and those which are manipulable variables. The output of the school is student achievement in cognitive areas, non-cognitive functioning such as motivation and attitudes, and social functioning such as relations with other people.



In this framework an indicator of a school's performance is taken to mean the difference between the expected (i.e. accounting for initial pupil's achievement and external factors) and the actual level of output. Focus is on performance for grades three and six and on the gain in achievement from grades one to three and from grades three to six.

The Performance Indicators in Education system was developed with several uses in mind. In the first place it will help policy-makers to identify more objectively the high- and low-performing programmes. It will contribute to the identification of educational needs to determine appropriate approaches to meet the needs and to evaluate the results. A first study was done in 1967. In 1974 a second study was published, this time focused on reading and mathematics only. In 1975 three studies were done on the relationships between teacher characteristics and achievement, between expenditure and achievement, and between student variables. Few consistent findings were delivered.

From this brief review of the full reports presented here, it appears that there are a range of socio-cultural factors each having an influence on learning results and each being difficult to manipulate through the means of educational policy.

VI. DIVIDING CONSUMPTION AND PRODUCTION

This section presents two systems which divide production and consumption activities in different ways. The first is a study by the Urban Institute: *Social Indicator Models for Urban Policy – Five Specific Applications* (in Baker, 1978). The study attempts to develop social indicators, with specific interest in a conceptual model classifying social indicators and explicitly structuring the process of service or goods delivery. Five areas of social concern were used, of which "learning and education" was one.

In "learning and education", the application of education indicators involved a distinction between the processes of learning and the processes of providing education. Learning was viewed as "a consumption activity by which individuals acquire skills and knowledge", some of which are obtained in formalised educational settings. Education was viewed as a "production activity in which there is a formalised attempt to impart certain skills and knowledge to those participating".

The second study, done by the Public Services Laboratory in 1972 (in Baker, 1978), explored educational outcomes and their policy applications, in an attempt to provide accountability. It argued that "education" as a product has both consumption and investment characteristics. Product consumption indicators are divided into "quantity" indicators (numbers in the education system) and "quality" indicators (measures of change in competence such as achievement attitudes).

Investment indicators shed light on the effect of education on income and employment. Secondary effects of investment and consumption include economic growth, consumer efficiency and citizenship values while tertiary effects are essentially society's long-term expectations of education.

I. Consumption	*II. Investment*
1. quantitative	Effects on:
2. qualitative	1. income, employment
	2. economic growth, consumer efficiency, citizenship values
	3. expectations in society of education

VII. COMPOUND INDICATORS

The indicator systems in the previous sections usually address relatively simple concepts, which are directly measurable. The following two proposals advocate compound or composite indicators for more abstract concepts. "Definition and selection of a composite indicator to represent aspects of education systems admits to the complexity of such systems. There are so many separate yet allied aspects to each of the major concepts of a system that a multivariate description provides the only way in which a manageable description can be realised. As relatively few composite indicators are needed to define an education system space, the benefits of simplicity of description and relative ease of comprehension are retained" (Johnstone, 1981).

Using compound concepts has the advantage of expressing the relationship between several aspects of the system as one concept.

Meuret (1987) proposes an "Educational System's Balance Sheet" so that decision-makers share a restricted system of reference indicators with the public. The topics of indicators are based on three guiding principles:

1. They must respond to public concerns.
2. They should use the progress made in strengthening evaluation techniques and the conceptual bases of evaluation. A distinction must be made between education as an activity and the results obtained. A level of schooling is an activity, not a result, while the number of graduates is a result. From this derives another distinction, that between *efficiency,* which relates activity to the available means, *effectiveness,* linking results to the activity, and *productivity,* which relates results to the means.
3. They must complete the balance sheet of the system through appraisal of its internal trends.

In addition to the proposed indicators, a diagnostic system must be aware of the opinions of the teachers, students and the public in general.

Using these principles, Meuret suggests 20 indicators across seven domains: costs, activities, results, efficiency, effectiveness, productivity, equality.

Coombs and Lüschen (1976) argue that the performance of an education system cannot be judged just by examining the output of the system. Only when output is examined in relation to goals, costs, demands and social needs can system performance be evaluated. They present four items: effectiveness, efficiency, responsiveness, and fidelity, which relate the output of an education system to the important educational, economic, political, and social factors which influence the formulation of educational policy. Effectiveness is the degree to which system output approaches an explicitly stated goal of the education system. Efficiency relates to a favourable ratio between costs and product (the unit cost for producing qualified doctors is an example of a measure of efficiency). Responsiveness is the degree to which the educational output is in concert with social demand from groups such as the public, teachers, students and parents. Fidelity is the degree to which the educational output corresponds to the needs of the society. In complex societies the needs are highly differentiated and consensus about them will not always be available.

VIII. THE EDUCATION SYSTEM AS A WHOLE

The following paragraphs present a number of models which can be said to comprise a balanced representation of the education system. This choice is to some degree subjective, as models in this section have sometimes been presented elsewhere.

UNESCO (1976*b*) developed a system of education indicators in a three-tiered pyramid. The top is represented by compound indicators, comparable to warning signals on the dashboard in a car, the middle by education indicators, playing the role of measuring instruments, and the bottom of the pyramid contains all other information, e.g. raw statistical data. Compound indicators are highly selective and aggregated; they enable the specialist to identify quickly whether there are major problems, dysfunctions and imbalances in the education system, and, if so, in which parts or aspects of the system they are located. The system is not for international comparisons, but for diagnosing the system.

The education indicators are divided into three groups: inputs (students, teachers, facilities, monetary inputs), internal process (measures of efficiency such as teacher hours per pupil per year) and output (a balance of economic requirements and social demand). Hence such indicators as equality of opportunity. Similar themes are introduced by the OECD (1973*b*) and Olkinuora (1973).

Compound indicators are based on clusters of education indicators. They must be easy to interpret as warning signals. Therefore, compound indicators are constructed in such a way that they vary between 0 and 1, and for each compound indicator a "danger zone" is defined. The compound indicators are grouped into the following main areas:

- measures of overall availability and utilisation of educational services;
- measures of equality/disparity;
- measures related to financial aspects;
- measures of employment aspects.

Johnstone (1981) wrote a comprehensive book on education indicators for UNESCO proposing a general framework and a set of indicators. Johnstone distinguishes three parts in the education system: input, process and output, and stresses the congruence of the system with its environment. Educational input indicators relate the amount of a particular resource brought into the education system to the total amount available for distribution. They can also be indicators describing the aspirations held by society for education systems, with the preferences of the political sector, those who provide, being distinguished from the private sector, or those who directly consume. Educational process indicators describe the structure of the system, or the distribution of the inputs throughout the education system. Educational output indicators relate the amount of a particular outcome of education to the global amount of some similar characteristics, or describe society's perception of the results of the education system's functioning. Congruence indicators measure the congruence between the output of resources and skills, and the needs of the society into which the outputs flow. It is through these measures that the effects of the education system on a society are noted. The external efficiency of the system can thus be assessed.

Windham (1988) gives the following model for the measurement of effectiveness of education systems:

Determinants

Inputs	Processes
– Student characteristics	– Forms of instructional organisation
– Teacher characteristics	– Alternative technologies
– School characteristics	– Use of teacher and student time
– Instructional characteristics and equipment characteristics	
– Facilities characteristics	

Effects

Outputs	Outcomes
– Cognitive achievement	– Employment
– Improved manual skills	– Earnings
– Attitudinal changes	– Status
– Behavioural changes	– Attitudinal changes
	– Behavioural changes

The production process for education consists of four main parts: inputs, processes, outputs and outcomes. Outputs are the direct and immediate effects of the educational process. Outcomes of education are the less direct and less immediate results of the interaction of educational outputs with the social environment. They cover a wider range of individuals than the leavers of the education system, a longer time period, and a larger political-geographical reference area. Outcomes include such personal effects as lifetime employment and earnings, consumption behaviour, job- and life-satisfaction, increased complementarity of labour skills to capital, and personal behaviour and attitudes. The term "characteristics" in the input section refers to the availability of a resource, its nature and quality, and the manner and rate of utilisation.

In a critical review of the "The Condition of Education", Herriott (1979) presents still another model of the education system. This model is divided into six parts:

1. *Environment:* those objects and patterns of relationships that exist outside an education system at any level, but significantly influence it or are influenced by it.
2. *Structure:* the relatively stable patterns of social interaction which integrate the various elements of an education system at any level.
3. *Culture:* the shared sentiments, meanings and commitments which various actors (teachers, pupils, administrators) within an education system at any level attach to its goals, activities and problems.
4. *Output:* the knowledge, skills, and orientations required by the environment and possessed by students at the time they leave an education system at any level.
5. *Throughput:* the raw materials (basically the students in attendance) which an education system at any level operates upon in producing output for its environment.
6. *Input:* the materials (textbooks), personnel (teachers and administrators) and information (knowledge and values) imported by a system at any level from its environment and used to transform throughput into output.

Besides the "process" aspects (input-throughput-output) the sociological aspects of structure and culture are depicted here as well as the aspect of environment.

IX. THE SCHOOL AS THE UNIT OF ANALYSIS

There are many indicator models in which the school is the unit of analysis. This approach seems to be increasingly favoured because of the popularity of the effective school movement and the attention now being given to the schooling process and the ways in which effective schooling can be promoted.

Garms (1987) attempts to develop the concept of merit schools which should be rewarded financially with non-merit schools being given mandatory help. Garms presents three possible models of defining merit schools, of which he prefers the last. In the first, one single criterion is measured: performance on a standardized test. The average achievement on this test for a school is predicted through knowledge of the socio-economic status of its students. The expected average achievement on the test for a school, predicted on the average socio-economic status of the students, is compared with the actual achievement in the school. Better actual achievement than predicted will be the basis for some reward to the school. The second method is similar, except that it uses

multiple outputs. The author does not specify the additional outputs to be examined. Both methods have shortcomings: they are too complex for the layman, the same variables must be used for all schools, and not all causal factors are known. In the third method, a number of output measures are used and weighted, some are mandatory, others are optional. This measure is much simpler and more flexible. The accent is on progression: a wide variety of indicators is used, the weights of the different indicators are different and known in advance, and a number of indicators are optional.

In California, a project called PACE (Policy Analysis for California Education) is underway. In 1984, California initiated a "merit schools plan". The first step was to define a set of performance indicators by which the state will rate each school and make the results public. The next step is for the state to award money to high-scoring schools. Therefore weights must be assigned to the different measures to provide an overall rating. The initial list contained 37 performance indicators, of which eight were measures of enrolment in advanced academic subjects, 19 were based on test scores, plus 10 other measures (Stern, 1987). Statewide targets for every indicator were given for the next years. The enrolment measures include: percentage of students meeting special requirements and percentage of students completing at least a number of years in important courses (English, maths, science). The test scores are California Assessment Program (CAP) tests for reading and maths at grades 3, 6, 8 and 12; Scholastic Aptitude Test (SAT); and advanced placement exams. Other indicators of success include writing assignments (how many write more than one essay per week at different levels), homework (more than one hour per day), dropout and attendance rates (percentage present on a particular day).

Benveniste (1987) offers a more comprehensive design for a school accountability system. He suggests that measures which provide direct information to teachers, parents and administrators be incentives for improvement. Developing this idea further, Benveniste argues that accountability schemes should focus on inputs, process or outputs. If there is a well-defined goal and a strong theory on how to achieve it, if the process is understood and what goes into the process is known, then a rigorous accountability system could be developed. Output measures work best when there is a consensus on what is to be achieved, when the measures are valid and reliable and when they have few unforeseen consequences. They do not work well when several simultaneous goals are pursued and some of these goals cannot easily be measured while others can. Measuring and controlling outputs amounts to saying "we want you to place large percentages of your students in college, but we do not care how you do it". Measuring and controlling the process is saying "do it this way" (with the assumption that a "best" way of performing the task is known and can be advised). Looking at input measures is a way of saying "here are the resources: are these adequate to achieve results?".

The focus is on schools as the relevant performance unit. Schools should be ranked in five categories: below state requirements, meets state requirements, improving school, research school and mentor school. The ratings should be sensitive to variables such as urban/rural, student turnover, socio-economic status and linguistic composition of student body. Additional resources would be provided by state agencies (top-down incentives), by reporting to local parents groups, school boards and others, and in this way schools might make better decisions (bottom-up accountability).

The measures must be parsimonious. They are of six kinds:

1. *Teacher preparation and achievement* (input);
2. *Teacher time* (process);
3. *Student learning time* (process);
4. *Order and consistency* (process e.g. truancy, co-operative behaviour);
5. *Parent and community support* (input e.g. volunteer hours, financial contribution);
6. *Student ability and outcome* (input and output e.g. test score).

In the Netherlands, Scheerens *et al.* (1988) propose a system for continuous follow-up of a number of variables of primary and secondary schools. Besides information on resources and input, indicators are needed on output, and the supply and organisation of the educational process. They suggest the following model:

Context:

- Size of school;
- Age of school building;
- Location;
- Level of service of schools;
- Denomination;
- Level of urbanisation.

Input:

- Pupils: distribution by social origin, scores on achievement tests, and ethnicity;
- Teachers: percentage older than 50, women and part-time;
- Resources: average expenditure per pupil.

Process:

- Achievement-orientation of school: percentage failing, selection of entrants, use of achievement norms, participation in achievement testing;
- Effective learning time: net number of lessons;
- Structure of school process: regular evaluation by teachers of school process, same methods in whole school, registration of the school career of students, presence of teachers in school outside teaching hours;
- Curriculum content: methods used.

Output:

- Percentage succeeding in examination;
- Percentage doing re-examination;
- Percentage of dropouts;
- Percentage upstream;
- Percentage going through after examination;
- Percentage downstream;
- Subjects chosen by pupils;
- Deviation score from central examination.

In the higher education sector in the Netherlands, an indicator system focusing on individual institutions is emerging in conjunction with the government's "Higher Education and Research Plan" which gives a number of education indicators. These indicators

will play a role in the dialogue that is to take place between the government and the institutions of higher education within the framework of the government's policy of steering from a distance. The government wants to reduce its control over the internal workings of institutions. At the same time, the higher education system's outputs should be more efficiently assessed in light of their response to society's functioning. Higher education institutions must become more autonomous, the quality of education and research must be improved and the institutions have to account for that. It is expected that indicators will assist in the allocation of resources – well-performing institutions will get more, weaker ones less. It is also expected that indicators can be used to improve the operationalisation of the institutional goals, thus leading to a more effective production process. In the existing situation the government has only poor knowledge of the spending, output and quality of the tasks exercised.

The indicators themselves are in the development phase and are subject to discussion because they can often be interpreted and evaluated differently, depending on goals and values. This attempt hopes to arrive at meaningful indicators of sufficient validity and reliability. The indicators considered important for education are (apart from a number of indicators concerning research and personnel):

- increase in numbers of students enrolled for the first time;
- ratio of total numbers of students to the first enrolled;
- graduation rate of first year;
- change in graduation rate of first year before and after curriculum reform;
- same as former but covering complete study period (In't Veld, Spee and Tseng, 1987; see also OECD/IMHE, 1988).

An interesting conceptual framework for the redesign of the United States' statistical data system for education is presented in *Alternatives for a National Data System on Elementary and Secondary Education*, by Hall *et al.* (1985). They argue that a conceptual framework for the assessment of educational quality is required to guide choices about such issues as the information to collect or the form in which it should be made available. Such a framework will also lead to a more cohesive system, i.e. an information system which allows relevant linkages between its components so that better assessments can be made of the effects of changing one part. A conceptual framework enables the setting of priorities within the system so as to achieve maximum benefit from the available resources.

The conceptual model they advance is based on the premise that an education system is an organisation which converts resources into educational services for pupils and its success is predicated upon its outcomes. Society intends the system to prepare individuals for work, for political participation, and for family life. Thus, creating an adequate national information system requires a more comprehensive definition of educational quality and a more systematic framework to describe an education system. Schools are confronted with wide variations in the educational difficulties they face. Assessing and improving the quality of education requires considering student achievement in relation to these difficulties and modifying the educational efforts made through the process of schooling. Merely focusing on achievement means to praise those educational units that draw students from educationally advantaged backgrounds, i.e. schools and districts of high socio-economic levels; this is a drawback to those who educate disadvantaged students.

The understanding of educational processes implies reliance upon a conceptual model. The model is a prerequisite to any understanding of the effectiveness and quality of schooling. The conceptual model focuses on the school because it is at the school level that educational activities take place and that pupils participate in them. This does not mean that all primary educational decisions and actions take place or should take place at the level of the school. Decisions of central importance about resource allocation, educational goals, and educational activities and programmes are taken at the state or district level. This does mean that once these decisions have been taken and constraints are imposed, much of the process of implementation – i.e. applying the allocated resources to create educational experiences which reach the desired outcomes – takes place at the school. Thus, any framework for constructing a comprehensive description of education systems should begin with the school and expand from it to the other levels of the system.

Schools and the communities they serve differ in several important ways:

- family and community environment;
- educational difficulty;
- resources;
- goals;
- process;
- outcomes.

It is vital to describe each of these differences in a cohesive fashion, within a coherent conceptual frame, as well as to attempt to sort out the reasons for differential outcomes against the structure of their origins. In the framework, teaching activities are distinguished from pupils' exposure and participation in the resulting educational activities. Quality of schooling resides in the process of schooling. Schools differ in the educational experiences they provide and in their success in gaining pupils' participation. These experiences are the services that a school provides to pupils and their parents. A good school is one which provides effective learning experiences or services actually leading to set goals and which accomplishes pupils' participation in those educational activities. The quality of these activities is then reflected in the pupils' achievements. But the knowledge and skills a pupil starts with will limit the outcomes of even the most effective schooling, and pupils who do not bring to school attention, perseverance, and the ability to digest rapidly the instruction they receive will gain less from any school experience. Thus, two education systems with similar goals, but different achievement levels, can only be assumed to differ in quality if they also face similar educational difficulties with their pupils.

In order to assess the quality of the education system, the data gathered must be comprehensive and linkages between the aspects of schooling must be revealed. Some of these linkages are:

i) Quality and effectiveness of education systems. It is the discrepancy between the desired and the actual outcomes which diagnoses the effectiveness of the system. For quality goals, background difficulties, resources and outcomes must be linked.

ii) Productivity and efficiency. To the extent that an education system is of the same quality or effectiveness as another, but has fewer resources to accomplish it, it is more efficient or productive.

X. CONCLUSION

In the last 15 years there have been numerous proposals to develop systems of education indicators. There is some difference of approach between these proposals, but there is also a large degree of overlap. The shared basis of these approaches is an input-process-output model. The elements that belong to these three parts of the education indicator systems are not always the same since the distinction between input and process, and between process and output, is not always made in the same way.

The role of the environment in an indicator system is also disputed. Some people regard environment as an explanatory variable. That is of importance in so far as it has an influence on the outcomes of the educational process. Others are looking more to the effect that education has on the environment. For them, education is not a goal in itself, but rather has a function to fulfill in society. And it is predominantly in this respect that the results of education must really be evaluated. For that reason the output of education can be split up into two parts: the direct results and the longer-term results. These two parts are called output and outcome, e.g. by Windham (1988) and US Bureau of the Census (1980).

Developments in the environment as explanatory variables are certainly important for explaining certain movements of indicators, but this does not mean that environment must be included in a system of indicators. The heading environment includes an enormous amount of information that in itself says nothing about the performance of the education system.

Sometimes the concept of an indicator is limited to the output (and outcome) of a system, and input and process are included only in so far as they are related to the output, so that compound measures of efficiency, productivity and effectiveness can be calculated. Input and process variables in themselves say nothing about the output of the system and can lead to long lists of indicators since their explaining power is often very limited. Output indicators are generally seen as more important than other types, and research has contributed to establishing relationships between input and process indicators on the one hand, and output indicators on the other. Compound indicators, which have up to now been seldom used, provide a useful means of limiting the number of indicators necessary for a meaningful set. These arguments do not combine to devalue completely simple indicators in the input, process and context domains.

A related issue is that typologies are often one-sided. Researchers belonging to the effective school movement focus on process indicators, economists on the relationship between input and output, political scientists are keen on steering possibilities, and sociologists look at the environment. It is important to find a balanced approach embodying these different outlooks so as to properly portray the complex reality of education.

Indicator systems are used at different levels: the national level, state level, the urban level and at the level of the school or the university. Because the users of the different levels will have different possibilities to steer, relative to their "degrees of freedom", they will have different information needs. For reasons of efficiency with regard to data-compiling and co-ordination in general, it may seem sensible to include the different levels into one comprehensive system, but the uses of such a system will differ according to the level of analysis.

Most systems restrict themselves to the "school" system and there is relatively little attention given to adult education. Though it is possible to distinguish relatively autono-

mous systems for schools, higher education and adult education, it would be useful to build these three systems into one broader education indicator system.

A first step towards such a system would be the development of a glossary of concepts so that gradually a common language will be used for the de /elopment, implementation and use of education indicators. This will ensure that the results of developmental work can be readily shared and increase the capacity for co-operative work in this area.

Thus the organising frame of the OECD/CERI International Education Indicators Project is seen to be simple and somewhat in need of clarification. The conceptual models which have preceded its development are important in placing the frame and the work of the Project in context.

Annex 1

CONCEPTUAL MODELS IN SYSTEMATIC ORDER

1. *Social indicators of education*

   OECD, 1982
   US Bureau of the Census, 1980
   United Nations (Stone), 1975
   Delors, 1971

2. *The inductive approach*

   Technical Committee in Education Statistics in Canada, 1988
   Quebec, Ministry of Education, 1986
   UNESCO, 1976*b*

   Japan, 1969 (in Yamamoto, 1987)

3. *Policy goals*

   Netherlands Central Bureau of Statistics, 1988
   OECD, 1973*a*
   Conrad, 1973
   Olkinuora, 1973

4. *Manipulable input variables*

   Council of Chief State School Officers, 1987
   United States (OECD, 1987)
   National Center for Education Statistics, 1985
   BSPE, 1972 (in Baker, 1978)
   The Urban Institute I, 1972 (in Baker, 1978)

5. *Dividing consumption and production*

   The Urban Institute II, 1975 (in Baker, 1978)
   The Public Services Laboratory, 1972 (in Baker, 1978)

6. *Compound indicators*

   Meuret, 1987
   Coombs and Lüschen, 1976

7. *The education system as a whole*

   Windham, 1988
   Johnstone, 1981
   Herriott, 1979
   UNESCO, 1976*a*

8. *The school as unit of analysis*

Scheerens *et al.*, 1988
Garms, 1987
Stern, 1987
Benveniste, 1987
In't Veld *et al.*, 1987
Hall *et al.*, 1985

Annex 2

CONCEPTUAL MODELS IN CHRONOLOGICAL ORDER

1969 Japan (2)*

1971 Delors (1)

1972 The Public Services Laboratory (5)
The Urban Institute I (4)
BSPE (4)

1973 Conrad (3)
OECD (3)
Olkinuora (3)

1975 The Urban Institute II (5)
Stone, in United Nations (1)

1976 Coombs and Lüschen (6)
UNESCO (7)
UNESCO (2)

1979 Herriott (7)

1980 US Bureau of the Census (1)

1981 Johnstone (7)

1982 OECD (1)

1985 Hall *et al.* (8)
National Center for Education Statistics (4)

1986 Quebec (2)

1987 Garms (8)
Stern (8)
Benveniste (8)
In't Veld *et al.* (8)
Meuret (6)
Council of Chief State School Officers (4)
United States (OECD) (4)

1988 Scheerens *et al.* (8)
Netherlands Central Bureau of Statistics (3)
Windham (7)
Technical Committee in Education Statistics of Canada (2)

---

* The figures in brackets refer to the clusters of Annex 1.

BIBLIOGRAPHY

Averch, H.A., S.J. Caroll, T.S. Donaldson, H.J. Kiesling, and J. Pincus (1971), *How Effective is Schooling? A critical review and synthesis of research findings,* Final report to the President's commission on school finance, Santa Monica, California, The Rand Corporation.

Baker, H.R. (1978), *Educational Indicators of Development: education and socio-economic development,* Monograph III, Saskatoon: Saskatchewan University.

Benveniste, G. (1987), "The Design of School Accountability Systems", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Conrad, Ch. (1973), *Die quantitative Erfassung qualitativer Aspekte von Schulsystemen,* Berlin: Technische Universität.

Coombs, F.S., and G. Lüschen (1976), "System performance and policymaking in West European education: Effectiveness, Efficiency, Responsiveness, and Fidelity", *International Review of Education,* 22, No 2: 133-153.

Council of Chief State School Officers (1987), "Education Assessment Program", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Delors, J. (1971), *Les indicateurs sociaux,* Paris: Futuribles.

Duncan, O.D. (1969), *Toward social reporting: Next steps,* New York: Russell Sage Foundation.

Garms, W.I. (1987), "Merit Schools for Florida, a concept paper", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Hall, H., R.M. Jaeger, C.P. Kearney, and D.E. Wiley (1985), *Alternatives for a National Data System on Elementary and Secondary Education,* A report prepared for the Office of Educational Research and Improvement, US Department of Education.

Herriott, R.E. (1979), "Some Observations on The Condition of Education" (1976 Edition), *Social Indicators Research,* 6: 181-196.

Johnstone, J.N. (1981), *Indicators of Education Systems,* London: Kogan Page; and Paris: UNESCO.

Land, K.C. (1975), "Theories, models and indicators of social change", *International Social Science Journal,* Vol. 27, No. 1: 8-37.

Mare, R.D. (1981), "Trends in schooling: demograghy performance and organisation", *Annals of the American Academy of Political and Social Science,* January, 96-122.

Meuret, D. (1987), "Some Indicators to Aid in the Global Evaluation of the Educational System", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Ministry of Education (1988), *Educational Indicators for the Elementary and Secondary Levels,* Government of Quebec.

National Center for Education Statistics (1985), *Indicators of Education, Status and Trends,* Washington, D.C.: US Department of Education.

National Center for Education Statistics (1987*a*), *The Condition of Education: A Statistical Report,* Washington, D.C.: US Department of Education.

National Center for Education Statistics (1987*b*), *Elementary and Secondary Education Indicators in Brief,* Washington, D.C.: US Department of Education.

Netherlands Central Bureau of Statistics (1988), *Educational Indicators for the Basisvorming* (comprehensive schools in secondary education, first stage), Voorburg: CBS.

Oakes, J. (1986), *Educational Indicators, a guide for policymakers,* Center for Policy Research in Education, OPE-01.

OECD (1973*a*), "A Framework for Educational Indicators to Guide Government Decisions", Paris: OECD (document for general distribution).

OECD (1973*b*), *Indicators of Performance of Educational Systems,* Paris: OECD.

OECD (1982), *The OECD List of Social Indicators,* Paris: OECD.

OECD (1987), "International Conference on Educational Indicators, Proposal by the United States", Paris: OECD (internal document).

OECD/CERI (1988), "International Educational Indicators: A working plan", Paris: OECD (internal document).

OECD/IMHE (1988), "Performance Indicators in Higher Education: A Study of their Development and Use in 15 OECD Countries", Paris: OECD (document for general distribution).

Olkinuora, E. (1973), "On the problem of developing educational indicators", *Acta Sociologica,* Vol. 16, No. 4: 284-302.

Saalfeld, B.F. (1973), *A Handbook of Educational Indicators. A Staff Report,* Salem, Oregon: Educational Coordinating Council.

Scheerens, J., W.G.R. Stoel, C.J.A.J. Vermeulen and W.J. Pelgrum (1988), *The Feasibility of a System of Educational Indicators for Primary and Secondary Education,* Enschede: University of Twente.

Stern, D. (1987), "Toward a Statewide System for Public School Accountability: A Report from California", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Technical Committee in Education Statistics in Canada (Primary and Secondary Education) (1988), *Definitions, Timing and Collection of Indicators adopted by the Steering Committee,* 1988-06-15, Ottawa: Statistics Canada.

UNESCO (1976*a*), *Towards a System of Educational Indicators,* Division of educational policy and planning, Paris: UNESCO.

UNESCO (1976*b*), "Current Social Indicators in the Field of Education", UNESCO/CES/AC.23/21, Paris: UNESCO.

United Nations (1975), *Towards a System of Social and Demographic Statistics,* New York: United Nations.

US Bureau of the Census (1980), *Social Indicators III, Selected data on social conditions and trends in the United States,* Washington: US Government Printing Office.

Veld, R.J. in't, A.J. Spee, and H.S.S. Tseng (1987), "Performance Indicators in Higher Education in the Netherlands", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Wainer, H., P.W. Holland, S. Swinton, and M.H. Wang (1985), "On State Education Statistics", *Journal of Education Statistics,* Vol. 10, No. 4: 293-325.

Windham, D.M. (1988), "Effectiveness Indicators in the Economic Analysis of Educational Activities", *International Journal of Educational Research,* Vol. 12, No. 6.

Yamamoto, S. (1987), "Educational Indicators in Japan", Paper presented at the Washington conference on education indicators, 3-6 November 1987.

Chapter 3

PROCESS INDICATORS OF SCHOOL FUNCTIONING

by

Jaap Scheerens
Professor, University of Twente
the Netherlands

I. INTRODUCTION

Education indicators are statistics that allow value judgements to be made about key aspects of the functioning of education systems. To emphasize their evaluative nature, the term "performance indicator" is frequently used.

Included in this definition of education indicators are:

- the notion that we are dealing with measurable characteristics of education systems;
- the aspiration to measure "key aspects", be it only to provide an "at a glance profile of current conditions" (Nuttall, Chapter 1) rather than in-depth description;
- the requirement that indicators show something of the quality of schooling, which implies that indicators are statistics that have a reference point (or standard) against which value judgements can be made.

Usually, policy-making at the national level is seen as the major source of application of indicators, and indicator systems are used as policy information systems. This view of the application of indicators should be enlarged, however, since consumers and "third parties", such as private industry, are also seen as users of the information that indicator systems provide. Likewise, the education system at local administrative level and even individual schools could also use indicators to support policy-making, i.e. using indicator systems as management information systems.

During the past decade collections of indicators, usually referred to as indicator systems, have been proposed and sub-sets of these have been used. Van Herpen gives a comprehensive overview of what he calls "conceptual models of education indicators" (Chapter 2). For our purpose, it is sufficient to discern some major developments in these various approaches to conceptualising education indicator systems.

Education indicators have their origins in the economic and social indicator movements. "Social indicators of education" describe educational aspects of the population, whereas education indicators describe the performance of the education system (see

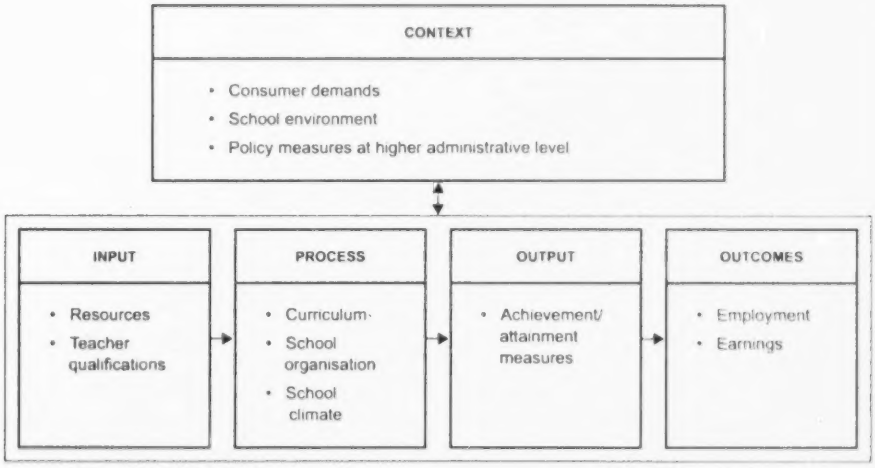
Chapter 2, Section II). In the development of education indicators, there have been three discernible trends.

First there was a transition from descriptive statistics to measuring performance, or, more generally, a shift towards statistics of evaluative importance.

When we look at developments in education indicators at the National Center for Statistics of the US Department of Education we can discern a second trend. At first this centre offered descriptive statistics on the state of the education system, including data on inputs and resources. Since 1982, "outcome" and "context" data were given a more prominent place, and in a recent proposal to redesign the education data system, "process" aspects of the functioning of education systems were also included (Stern, 1986; Teauber, 1987). This second trend is a movement towards more comprehensive indicator systems, first adding output measures and context measures to the more traditional measurement of inputs and resources, and secondly by a growing interest in "manipulative input factors" and process characteristics.

The third trend is somewhat related to the second, as far as the interest in process characteristics is concerned. Traditionally, indicator systems have concentrated on macro-level data, such as national illiteracy rates, the proportion of pupils who have passed their final secondary examinations, schools, etc. When we think of process indicators as referring to the procedures or techniques that determine the transition of inputs into outputs, interest in process indicators naturally leads to an interest in what goes on in schools. So, the third trend in developing education indicators and conceptualising indicator systems is to measure data at more than one aggregation level: national system, school, perhaps even the classroom (see Teauber, 1987; Scheerens *et al.*, 1988).

Figure I. **"Context–input–process–output–outcome" model of schooling**



What emerges from this brief overview of developments in the field of education indicators is the notion that a context-input-process-output model in Figure 1 is the best analytic scheme to systemise thinking on indicator systems.

This report focuses on the process component of this model, initially by a discussion of the specific place of process indicators within indicator systems. A review of the literature on school and teaching effectiveness will follow in an attempt to identify those school and teaching characteristics which are positively associated with educational achievement. From this review, a comprehensive model of effectiveness can be developed and used as a basis for a list of the most promising process indicators.

II. BASIC QUESTIONS RELATED TO PROCESS INDICATORS

The question arises whether process measures of school functioning fall within the definition of education indicators that has already been given. It is doubtful whether process measures as such can be used as a basis for judging the performance of an education system.

Would one, for instance, be willing to accept the degree to which a school or a school district used a specific curriculum as a trustworthy criterion for judging educational performance? Process measures relate to outcome measures as a means to an end and thus, it could be argued, using these as performance indicators would be goal displacement in a new form. To put it bluntly, process indicators could lead to evaluative conclusions of the kind "operation successful, patient deceased". The only way out of this would be the existence of empirically-supported causal models of educational performance, from which the importance of specific process measures could be deduced. Unfortunately no such established causal models exist. Generally, the variance in output measures that is accounted for by input and process measures is rather low. In the author's view, the only legitimate way to employ process indicators is always to link them to output indicators. Process indicators then have the function of offering hypothetical explanations on why certain schools, or school systems, do better than others. The notion that process indicators derive their value from their relationship with output indicators forms the basis of this report; namely to select process indicators by examining the findings of research on school effectiveness.

Process indicators generally refer to characteristics of education systems that can be manipulated. Adding process measures therefore enhances the policy relevance of indicator systems.

The question of measurement is another point. Generally, indicators are thought of as quantitative "low inference" measures. Process characteristics such as school climate, educational leadership and opportunity to learn are rather broad characteristics; their operationalisation and quantification are neither straightforward nor "low inference". Further, elaborate procedures sometimes have to be used to collect such data. Remaining strictly within the domain of indicators, this problem can be solved only by using "proxy" measures of process characteristics, for instance by measuring instructional time during school days in a year as a proxy for the time students spend on task-related activities, and by omitting all process variables that are not amenable to these kinds of approximations. Another way of tackling this problem is to use in-depth studies (surveys or observational studies) that are connected to regular indicator systems.

III. RESEARCH ON SCHOOL EFFECTIVENESS AND THE IDENTIFICATION OF PROCESS INDICATORS

As has already been stated the most likely rationale for selecting process indicators is to choose those variables that are manipulative predictors of school output. Research literature on school effectiveness can be used as a source for identifying promising process variables. Generally speaking, research on school effectiveness is aimed at discovering school characteristics that are positively associated with school output, usually measured as students' achievement. Various research traditions can be subsumed under this heading, including (in)equality of education (sociological), educational production functions (economic), school improvement and effective schools, and teacher and instructional effectiveness (psychological). Apart from these, more theoretical and analytic contributions from organisational science and micro-economic theory of public-sector organisations can also be sources of inspiration in selecting process indicators (see Cameron and Whetten, 1983; and Niskanen, 1971).

So what are the main findings of research on school effectiveness? What proposals for process indicators can be generated from these findings?

Research on (in)equality in education and school effects

The Coleman report (Coleman *et al.*, 1966) on the Equal Educational Opportunity Survey should be seen as the impressive starting point for research on school effectiveness. Although the major thrust of the Coleman survey was to investigate (in)equality in education, it also became quite famous for its supposed negative conclusions on the influence of the school on educational achievement. Coleman *et al.* found that schools accounted for approximately 10 per cent of the variance in pupil achievement, after statistical adjustments had been made for the influence of background characteristics of pupils. The significance of this finding for judging the importance of process indicators on school functioning is that all feasible process indicators one could think of would not account for more than 10 per cent of the variance in pupil achievement.

However, depending on the achievement measure that is used and the heterogeneity versus the homogeneity factor of a school system in a particular country, later research has found a higher percentage explained by schools. Coleman's 10 per cent is not exceptional when considering more recent findings (cf. Purkey and Smith, 1983; Bosker and Scheerens, 1989). Other large-scale studies replicated Coleman's findings in their pessimistic conclusions on the importance of schooling and its possibilities to lower educational inequality (Jencks *et al.*, 1972, 1979; Hauser *et al.* 1976; Thorndike, 1973). Specific school characteristics that were measured in these studies were mainly resources and material inputs such as the age of the school building, per-pupil expenditure and the number of books in the library, although some measures of teacher attitudes and instances of classroom management were used. The significance of this first generation of research on school effectiveness for the issue of education indicators can be summarised in three points:

a) School process variables account for relatively little variance in educational achievement.

b) Resources and "material" inputs are not very promising in explaining school output, though this would not necessarily imply that they should not be included

in indicator systems. In school systems in developing countries, for instance, they might still be of great importance.

c) Pupil background characteristics such as socio-economic status or race should be used to adjust raw output measures to arrive at fair and valid performance indicators and to allow for an unbiased interpretation of the influence of process characteristics on the functioning of schools.

Research on educational production functions

The economic approach to school effectiveness is concerned with the cost of inputs and with the question of which inputs lead to more output. Stated in more abstract terms, knowledge about stable relationships between input and output variance is sought in order to specify a function that could express the effects more inputs would have on output. This "effectiveness" research is known both as input-output analysis and as research on educational production functions. In fact, this type of research is very similar to other types of research on educational effectiveness in that the relationship between school characteristics and achievement is investigated, while adjusting for background characteristics of pupils such as level of intelligence and socio-economic status. The characteristic that sets this research tradition apart is the choice of a particular category of inputs that are readily expressed in monetary terms, such as teacher salary, teacher experience, teacher/pupil ratio, teacher qualifications, and per-pupil expenditure.

The results of this type of research on effectiveness are rather disappointing. Research reviews such as those of Mosteller and Moynihan (1972), Averch *et al.* (1974), Glasman and Biniaminov (1981), Hanushek (1979, 1986) agree upon the inconsistency of research findings and the rather small effect of the input variables concerned.

For example, Hanushek (1986) reviewed 147 input-output studies with inputs being teacher/pupil ratio, teacher qualification, teacher experience, teacher salary and per-pupil expenditure. Of these, only teacher experience showed any consistency with 30 per cent of the estimated coefficients being statistically significant. Hanushek's overall conclusion is that educational expenditure is not consistently related to achievement. He suggests that it would take greater variation in inputs to expect important effects. So, for instance, in most countries where teacher salaries are strictly regulated and rather uniform, if a system of "merit pay" would be introduced, the variance in teacher salaries would be expected to increase and significant effects in pupil achievement might be revealed.

Although this research tradition does not focus on "process" measures and notwithstanding the primitive nature of the production functions and the assumption that treating the whole area of educational technology is a black box, some of the input variables could be considered for inclusion in indicator systems. Judging from Hanushek's research synthesis, teacher experience would be the most likely candidate. Yet, when one wishes to construct education indicators for international comparison, it could be wise to include variables such as per-pupil expenditure and teacher/pupil ratio, as these might show significant variance between countries.

Research on effective schools

After the Coleman report, a second wave of research on school effectiveness came into being. Its pioneer studies can be seen as a reaction against Coleman's negative conclusions. As titles such as "Schools can make a difference" (Brookover *et al.*, 1979) and "School matters" (Mortimore *et al.*, 1988) show, refuting the message of the Coleman report has been, and still is, an important motive for this more recent research. The most important characteristic that distinguishes the research on effective schools from earlier work is that the black box of what happens within schools is opened and school variables are revealed to include school organisation, school culture and educational technology. Several types of research on effective schools can be distinguished.

First, there is a series of studies where exceptionally effective schools were identified and described as case studies, sometimes also comparing them to ineffective schools in order to discover which of the school characteristics were responsible for the superior results. Examples of these studies are those by Lezotte, Edmonds and Ratner (1974) and Weber (1971). School characteristics that were consistently associated with positive achievement were:

- a safe and orderly school climate;
- high expectations of pupils' achievement;
- educational leadership (i.e. a school leader who is actively involved in developing and monitoring educational activities, and who is more than merely an administrator);
- frequent evaluation of pupils' progress;
- clear objectives concerning basic skills;
- a co-operative atmosphere among the teaching staff.

A second type of studies of effective schools arose because of the rather surprising fact that despite the relatively weak research basis for the above-mentioned studies, these findings were almost immediately used as a guiding principle for school improvement programmes. Although few of these school improvement programmes have been rigorously evaluated, those that have been show positive programme effects (e.g. Miller *et al.*, 1985; Achilles and Lintz, 1986; McCormack-Larkin, 1985). It should be noted that these evaluations of school improvement programmes show the overall effect of multiple changes in the functioning of schools, including better educational leadership, building an achievement-oriented morale among pupils, stimulating professional attitudes and co-operation among teaching staff, and applying structured teaching methods that include frequent evaluation, feedback and reinforcement.

A third type of research on effective schools comprises a blending of the approaches of the earlier large-scale input-output effectiveness studies with the more recent interest in process characteristics of school functioning. Studies such as those of Brookover *et al.* (1979) and Mortimore *et al.* (1988) combine relatively large sample survey research with in-depth description of school processes. Moreover, the latter study is not exclusively involved with characteristics measured at school level but also includes aspects of instruction at classroom level. By way of illustration, the factors that Mortimore *et al.* found to be positively associated with pupil achievement are listed in Table 1.

It is quite obvious that ideally these effective school characteristics would require intensive data collection and high inference measurement. Some of them, however, can

Table 1. **Twelve key aspects of effectiveness**

1. *Purposeful leadership of the staff by the headteacher.*
   Key aspects: active involvement in school's work without exerting total control over the rest of the staff, record-keeping.
2. *The involvement of the deputy head.*
   Key aspect: sharing of responsibilities of head and deputy.
3. *The involvement of teachers.*
   Key aspects: active involvement in curriculum planning, participation in decision-making on school policy.
4. *Consistency amongst teachers.*
   Continuity in the teaching staff and consistency of teacher approach.
5. *Structured sessions.*
   Key aspect: teachers offer a strict framework within which pupils can work, yet allowing them some freedom.
6. *Intellectually challenging teaching.*
   Key aspects: asking of higher-order questions, enthusiastic approach, high expectations of pupils.
7. *Work-centred environment.*
   Key aspects: much content-related work and feedback – relatively little time spent on routine matters – and the continuity of work activities; a low level of noise.
8. *Limited focus within sessions.*
   A focus upon only one curriculum area in a lesson.
9. *Maximum communication between teachers and pupils.*
   A flexible approach, using a blend of individual, class and group communications – whole class teaching increased the amount of communication with all pupils in comparison to an approach where teachers devoted the majority of their time to speaking with individual pupils.
10. *Record-keeping.*
    Record-keeping by both headteachers and teachers.
11. *Parental involvement.*
    Key aspects: help in classrooms, educational visits, attendance at meetings, parents' reading to their children, access to books at home.
12. *Positive climate.*
    Key aspects: more praise than blame, enthusiastic attitude of teachers, friendly attitude of pupils – absence of graffiti around the schools.

Source: Mortimore *et al.* (1988).

be operationalised in terms of relatively uncomplicated scales or questionnaire items. Some examples are:

Educational leadership

- the amount of time headteachers spend on educational matters, as opposed to administrative and other tasks;
- whether headteachers do or do not discuss test results on pupils' progress with teachers;
- the amount of instructional issues on the agenda of staff meetings;

Achievement-oriented policy

- the amount of overt statements in official school documents that express an achievement-oriented emphasis in school policy;

Orderly and safe climate

- statistics on absenteeism, lesson drop-out and delinquency as instances of the degree of order in the school;
- ratings of school discipline by teachers and headteachers;

Clear objectives

- whether or not explicit school curricula, stating educational objectives and levels of achievement are available;

High expectations

- estimates by teachers and/or headteacher of the percentage of students that will complete their secondary schooling;
- student estimates of their further education;

Monitoring/evaluation of pupils' progress

- the frequency of the use of curriculum-specific tests at each grade level;
- the same with respect to standardized achievement tests;
- whether or not the school uses a (computerised) system to monitor pupils' progress at all grade levels;

Continuity and consensus among teachers

- the amount of changes in staff over a certain period;
- the presence or absence of subject-based communication platforms (secondary education);
- the degree of opportunity for communication and co-operation among staff.

Some characteristics associated with school effectiveness could be assessed by unobtrusive observations by researchers or inspectors such as the presence or absence of grafitti in and around the school and the noise level in school corridors. Other characteristics, mainly at classroom level, would require detailed observation and reporting by teachers and pupils. It would require creativity to find proxy measures for these effective school characteristics that would fit the requirement of easy, low-inference measurement. Some characteristics that might be readily available from school records include:

- staff turnover within a certain period;
- existence of a specific school curriculum;
- the presence or absence of procedures and records to monitor student progress;
- whether or not schools keep records on examination results and the percentage of students passing on to higher education.

Research on instructional effectiveness

It would be far beyond the scope of this report to attempt to give a comprehensive overview of the research literature on teacher effectiveness, effective classrooms, instructional effectiveness and the like. The distinguishing characteristic of this stream of educational research – compared with the various types of studies on school effectiveness, described earlier – is the fact that process characteristics of education are studied at the teacher or classroom level. So, when we are considering variables at this level that have been found to be associated with achievement, we are really delving into the primary process of schooling. We shall come back to the question of the place that these kinds of micro-level variables could have when considering indicator systems.

Where a detailed overview of the relevant research literature is out of the question, the next best solution in the search for the most promising process variables is to use already existing research reviews. Fortunately, many excellent reviews are available (e.g. Kyle, 1985; Brophy and Good, 1986). Also a growing amount of quantitative syntheses of research on educational productivity has become available (Walberg, 1984; Fraser *et al.*, 1987; Kulik and Kulik, 1982).

This review literature reveals a set of categories of variables that have repeatedly been shown to be positively associated with achievement (for a more detailed account of the selection of these variables, see Scheerens, 1989). These are:

i) Effective learning time or "time on task". Specific aspects of learning time are: the duration of the school day, school week and school year, whether or not pupils get homework assignments, the amount of the official duration of lessons that is actually spent on task-related work, absenteeism, truancy rates and reallocation of the total time that is available for instruction over school subjects. Issues of school discipline are also related to the amount of time that is effectively available for instruction.
Although interpreting the results of increased effective learning is straightforward, two points should be noted. First, it is obvious that extending the official school hours must at some point become counter-productive. Secondly, moderate increases in learning time have yielded only moderate effects on achievement (cf. Levin, 1988; Walberg, 1984).

ii) Structured or "direct" teaching. The general idea of structured or "direct" teaching is the application of frequent interventions to support the learning process. Examples of these include: stating educational objectives clearly, dividing the total subject matter that must be learned into relatively small units, providing a well-planned sequence of these units, allowing many opportunities for pupils to do excercises, giving cues and hints, frequent questioning and testing to monitor progress, and giving feedback.
Mastery learning is a didactic approach in which most of these principles are represented. Although direct teaching has been demonstrated to be particularly effective in primary education and for the teaching of basic skills, it has also been shown to work in secondary education and in the teaching of higher-order cognitive skills – though in a somewhat modified form: larger steps in subject matter presentation, more initiative for pupils (cf. Doyle, 1985; Collins and Stevens, 1982). Adaptive instruction, i.e. adapting instruction to pupil characteristics, in matters such as tempo and way of presentation, can be seen as a more individualised use of structured teaching.

iii) *Opportunity to learn or "content covered".* The essence of "opportunity to learn" is the correspondence between the subject matter that has been taught and the content of the tests that are used to measure achievement. As is to be expected, pupils do better when the subject matter is covered by test items.

iv) *Teacher attitudes and expectations.* As was shown in older literature on teacher research, the enthusiastic attitude of teachers is important. From more recent studies where effective teaching is compared to less effective teaching in inner-city schools, it seems important that teachers remain optimistic about the capabilities of their students; nothing seems worse than a defeatist attitude. This attitudinal factor of teacher functioning directly corresponds to the high expectations variable known from the literature on school effectiveness.

v) *Enhancing student motivation.* According to Walberg's (1984) research synthesis, the variable most strongly related to achievement is reinforcement. This variable is closely related to structured teaching, where frequent monitoring of progress and feedback are important. It appears that praise, or positive feedback, works far better than punishment (see Brophy and Good for a more detailed analysis).

vi) *The alterable curriculum of the home.* This category of variables – as it was named by Walberg (1984) – covers for the ways in which the home situation of students can affect school performance. Positive instances of this category include parental interest in what children do at school, reading to children at home and moderate television viewing.

IV. TOWARDS A MORE COMPREHENSIVE CONCEPTUALISATION OF SCHOOL EFFECTIVENESS

Before going into the question as to whether these instructional variables are amenable to relatively easy low-inference measurement, and more generally, what part they might play in indicator systems, an attempt will be made to integrate the results of the various types of studies on educational effectiveness into one general model. As already stated, the rationale that is being followed to identify and select useful process indicators is to look for those process variables that are regarded as manipulative predictors of educational achievement.

Thus, the development of a conceptual model of school effectiveness is of direct relevance in identifying prospective process indicators. Some of the building blocks of a comprehensive model of school effectiveness have already been presented: a context-input-process-output analytic framework and the most important results from various types of research on school effectiveness. To these, three "ingredients" must be added:

a) a multi-level framework which is most useful in specifying the input-process-output-context elements of school effectiveness models;

b) a perspective to analyse the relationship between context variables and process variables known as "contingency theory";

c) a perspective for the relationship between school-level organisational and managerial conditions and process characteristics of school functioning at the micro (i.e. classroom) level.

Multi-level framework

For the purpose of developing a conceptual model of school effectiveness it is most useful to think of the measurement of the different types of variables at the lowest level of aggregation at which they can be defined. Besides conceptual clarity, this has the advantage that available statistical techniques for multi-level analysis yield the most precise results. This implies that the context-input-process-output framework from which we start would need at least five levels:

- The measurement of output (achievement/attainment) and important background variables (socio-economic status, previous educational achievement, intelligence) at student level;
- The measurement of instructional process variables at teacher/classroom level;
- The measurement of organisational, curriculum and managerial process variables at school level;
- The measurement of material and financial inputs, also at school level;
- The measurement of context variables either at school level (i.e. school size) or regional level.

Contingency perspective

The general view of contingency theory is that it depends upon contextual characteristics, whether or not specific organisational structures or managerial processes will be effective (e.g. Mintzberg, 1979). At first glance, contingency theory seems at odds with research on school effectiveness, which is concerned with a set of school characteristics that is very robust in predicting effectiveness. On closer analysis of the research and literature on school effectiveness, it becomes clear that claims regarding the generalisation of effective predictors across contexts have only partly survived empirical tests (Firestone and Herriott, 1982; Teddlie *et al.*, 1987; Scheerens, Nanninga and Pelgrum, 1989). In fact, including contextual variables such as student-body composition, school type, or national educational context can be seen as a relatively new and very interesting development in research on school effectiveness. Some studies even try to establish interactions between various contextual conditions and performance (i.e. Friedkin and Necochea, 1988, who investigated the interactional effect, on student performance, of school size and socio-economic status of the student body).

Two applications of contingency thinking seem to be particularly relevant to the subject of school effectiveness. First, in organisational theory the notion of effectiveness as it is used in research literature on school effectiveness (i.e. productivity) is sometimes seen as just one particular type of effectiveness. Alternative effectiveness criteria are resource acquisition, stability and control in the functioning of the organisation and cohesion and morale among the organisation's members (Cameron and Whetten, 1983; Faerman and Quinn, 1985). Depending on contingency factors such as the degree of environmental uncertainty the organisation is faced with, or the stage of its development, the emphasis on each of these types of effectiveness criteria might shift. For instance, when a school is faced with important drops in enrolment, it is quite understandable that relatively more energy will be put into acquiring more pupils. And when a school, for whatever reason, has to merge with another one, a lot of attention will have to be given to reaching stability and control in the new organisation.

As research shows, schools differ in the degree to which achievement is emphasized in school policy and teacher expectations. Moreover, achievement orientation is generally found to be positively related to actual achievement. The organisational conceptualisation that recognises effectiveness criteria other than productivity and explains their relative influence in a particular setting by referring to contextual conditions helps to answer the question why some schools are more achievement-oriented than others.

The second instance of contingency thinking relevant to school effectiveness is recognising the importance of external incentives on achievement-oriented school policy.

The political will of a school to achieve is perhaps the most essential condition for actual school effectiveness. When higher administrative bodies, consumers, or other stake-holders also emphasize achievement or even reward schools for high achievement and "punish" others for low, this political will can even be seen as a malleable factor. In this respect a range of measures including output finance of schools, "privatising" schools, deregulation, voucher systems and publishing schools performance in local newspapers is relevant. Micro-economic theories on the efficiency of public-sector organisations (i.e. Niskanen, 1971; Breton and Wintrobe, 1982) explain how these mechanisms operate. For our purpose, we can put all these external mechanisms to stimulate schools to be effective under the heading "achievement incentives in the local and the larger school environment". At the local level we could look at the presence or absence of achievement standards for schools, clear objectives, evaluation systems and public records on school achievement. When comparing national education systems, one could include such measures as the presence or absence of assessment projects, the degree to which the inspectorate employs output evaluation, the development of consumerism on education and whether or not output finance is used.

Meso/micro relationships

In our earlier review of different types of research on educational effectiveness we discussed research which focuses on process characteristics at the school level and research on instructional effectiveness where processes at the teacher or classroom level are of central interest. Both research areas have yielded a list of the most promising process characteristics, i.e. school and instructional characteristics respectively. The relationship between these two categories is an important aspect of our envisaged comprehensive model of school effectiveness. The most straightforward way of seeing this meso (school level)/micro (classroom level) relationship is to assume that meso-level conditions facilitate micro-level conditions. This implies that instructional processes are seen as the most direct determinants of school learning and achievement, and that organisational and curricular conditions at school level are thought of as more indirect conditions of educational achievement.

When we examine the actual research outcomes on meso and micro conditions of educational achievement more closely, it is evident that some important variables are meaningful at both the school and teacher/classroom level:

- Structured teaching at classroom level can be stimulated by means of explicit curricular policy at school level (e.g. by using school development plans).
- Expectation of high student performance is essentially a variable defined at the teacher level, though its aggregate, an achievement-oriented school policy, may be taken as a whole that is more than the sum of its parts.

- Order in classrooms will be enhanced by an orderly atmosphere in other parts of the school building
- Frequent monitoring of pupils' progress will usually take place at classroom level, though this evaluation may be a result of a school evaluation policy and will benefit from instruments at this level, such as computerised school evaluation or a management information system.
- Opportunity to learn can be defined at the classroom level, but can also be seen as being enhanced by a school curriculum that is closely linked to the educational objectives that determine the contents of achievement tests.

Apart from these factors that can be defined at both school and classroom levels, we can discern a second class of conditions at school level that facilitate effective instruction at classroom level. Their successful operation is dependent on the organisation's superstructure (e.g. management, co-ordination structures) protecting the core production process against disturbances and external uncertainties (Thompson, 1967). Instructional leadership, the degree of collaborative planning and collegial relationships plus an active policy in recruiting students and acquiring resources are examples of the latter type of conditions that have received some support in research literature on school effectiveness.

Synthesis

The ingredients for a comprehensive model of school effectiveness are in place. To summarise they are:

- an analytic systems model recognising context, input, process and output variables;
- a multi-level framework discerning pupil, classroom, school and environmental characteristics;
- perspectives to view the inter-relationships between variables defined at different levels, most notably contingency theory and organisational conditions that facilitate the school's primary process;
- substantive findings from different types of research on educational effectivenes.

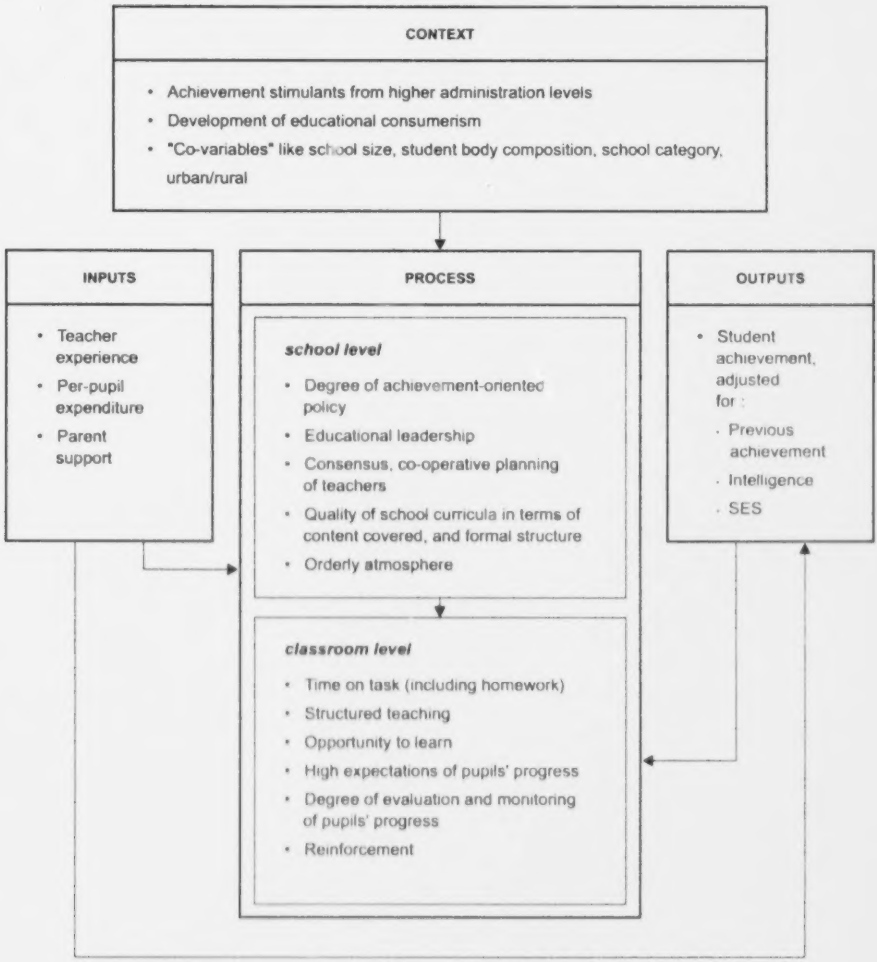
The hypotheses that have been stated on the inter-relationships of the various categories of variables are the extra dimension that this model offers over a mere listing of the most promising variables. Although empirical testing of this integral model, and developing the theoretical explanations of these hypothetical relationships are clearly beyond the scope of indicator development, its present tentative formulation might nevertheless be helpful in selecting those process variables that are most relevant in the causes of achievement differences between schools.

V. SOME PROBLEMS IN INTERPRETING THE RESULTS OF RESEARCH ON SCHOOL EFFECTIVENESS

Empirical basis

The equality of the research base that supports the list of factors presented in Figure 2 varies considerably among the individual variables. Some variables have been

Figure 2. **Integral model of school effectiveness**


CONTEXT
• Achievement stimulants from higher administration levels
• Development of educational consumerism
• "Co-variables" like school size, student body composition, school category, urban/rural
INPUTS
• Teacher experience
• Per-pupil expenditure
• Parent support
PROCESS
school level
• Degree of achievement-oriented policy
• Educational leadership
• Consensus, co-operative planning of teachers
• Quality of school curricula in terms of content covered, and formal structure
• Orderly atmosphere
classroom level
• Time on task (including homework)
• Structured teaching
• Opportunity to learn
• High expectations of pupils' progress
• Degree of evaluation and monitoring of pupils' progress
• Reinforcement
OUTPUTS
• Student achievement, adjusted for :
- Previous achievement
- Intelligence
- SES

shown to have a positive association with achievement in many studies, others, such as pupil expenditure, have been investigated many times but the results show no consistent pattern. Some other variables have only a relatively weak research base and their inclusion in the model is more a matter of conjecture than solid empirical evidence.

Table 2 gives an overview of the degree to which each of the factors in Figure 2 has been empirically supported.

Table 2. **Degree of empirical support for educational process variables associated with achievement**

Characteristic	Strong empirical basis	Moderate empirical basis	An as yet weak empirical basis	Mostly conjecture
Environmental incentives				x
Consumerism/parent involvement		x		
Teacher experience		x		
Per-pupil expenditure		x		
Achievement-oriented policy, high expectations		x		
Educational leadership			x	
Consensus, co-operative planning				x
Quality of curricula			x	
Evaluative potential			x	
Orderly climate			x	
Structured teaching	x			
Time on task	x			
Opportunity to learn	x			
Reinforcement	x			

Table 2 shows that there is not yet an empirically-supported comprehensive school effectiveness model. This should not preclude the use of the available research evidence as a reasonable desideratum to identify prospective process indicators. At the same time the available research base still leaves much uncertainty as to the status of certain factors, which also implies that other reviewers of the literature will probably produce somewhat varying lists of most important factors (see Purky and Smith, 1983; Bossert, 1988).

Narrow focus?

Although the use of educational attainment data (e.g. the proportion of an age cohort that passes the final examination of secondary school) is not at all uncommon in research on school effectiveness, most effectiveness studies have used achievement test data on a limited set of subjects, usually language and mathematics.

Given that national education systems, school districts and individual schools may differ in the degree to which they value these particular educational outputs, they will also consider differently the usefulness of school effectiveness models that rest upon this narrow definition of school output. A counter-argument to this possible source of normative relativity of research findings on school effectiveness is that the basic educational objectives that appear in tests and examinations do so exactly because they are very important.

Yet, it seems wise to keep an open mind to differences in normative contexts in which indicator systems, inspired by school effectiveness models, are to be applied.

"Small" effects

A problem in interpreting the results of research on school effectiveness is in how far the sizes of malleable educational characteristics are to be seen as educationally significant. It is common knowledge that these process characteristics generally explain relatively little variance in pupil achievement. Before going into this, it should be noted, first and foremost, that we are always dealing with relative school effects, that is, the degree to which schools differ among themselves in average achievement.

Defensive research titles such as *Schools can make a difference* and *School matters* seem to have been used as weapons against the straw-man argument that it does not matter whether or not children go to school. In a sample of about 30 studies on school effectiveness (Scheerens, 1989), the average variance between schools was about 12 per cent of the total (i.e. pupil-level) variance in achievement, ranging from 5 per cent to 40 per cent. (In this section, all figures on school effect sizes are based on achievement scores that have been adjusted for pupils' background characteristics.) This implies that specific school-level characteristics can only account for a certain proportion of the variance between schools, which means that they may not explain more than, say, 4 per cent of pupil-level variance. This tiny effect would, in a situation where variance between schools is 12 per cent, still mean that the variable in question explained one-third of all that specific school factors could explain in this particular situation.

In interpreting school effects we should bear in mind that up to 75 per cent of the variance in pupil achievement can be accounted for by factors such as intelligence and socio-economic factors, that are to a large extent beyond the reach of educative manipulation. This means that the total direct influence of schooling is relatively small as far as a pupil's overall achievement level is concerned.

Interpreting school effects in their proper perspective could benefit from more precise ways of expressing effect sizes (see Rutter, 1983; Bosker and Scheerens, 1989). A general approach to accomplishing this would be to attach some kind of societal value to score levels on the output variable. Purky and Smith (1983) give an example of this by translating a difference of two-thirds of a standard deviation in achievement between highest-scoring and lowest-scoring schools into a certain time interval (in their case one year) that expresses the degree the average pupil of the highest-scoring school is ahead of the average pupil of the lowest-scoring school.

The conclusion on the "small effects" in research on school effectiveness is that we should use more specific expressions of effects than merely "percentage of total variance explained" and by doing so we may discover that so-called "small effects" may still be of educational importance.

Conceptual and methodological problems in research on school effectiveness

Conceptual problems in interpreting the results of research on school effectiveness are closely related to the correlational nature of most studies and the "inductive" approach in building school effectiveness models. So, for instance, it is hard to say whether a variable such as "high expectations of pupils' progress" is to be seen as cause or effect of high achievement, when we have not experimentally manipulated the expectations variable. More generally, we may have some ideas now on what variables work in education, yet, we still have little knowledge on the causal mechanisms that explain the

correlations. The links with more general explanatory principles or theories are still relatively weakly developed (cf. Scheerens and Stoel, 1988).

Methodological weaknesses of research on school effectiveness have been intensively described in various review articles such as Ralph and Fennessey's (1983). "Small samples", "insufficient adjustment for important background characteristics of schooling", "reactive research arrangements" and "improper use of analysis techniques" are the main points of criticism.

Although these conceptual and methodological problems are certainly one more reason to treat the results of research cautiously, some counter-arguments should be considered too. As to the methodological criticism it should be remarked that results have shown a considerable robustness in the face of varying research settings and research approaches. An argument for the emergent school effectiveness model that is hard to neglect – though not very scientific – is its intuitive appeal. Some of the basic principles such as more success in specific kinds of school output when these outputs are actively strived for and are an explicit policy of the school, and better results when more time is invested, simply make sense. Finally, some attempts have been made to link the findings of research on school effectiveness to more general explanatory principles like learning theory, X-efficiency, and certain conceptualisations from organisation theory (Scheerens and Stoel, 1988).

Despite all cautionary notes that have been made in this section, it still seems permissible to use the research literature on school effectiveness as a best guess for proposing likely candidates for process indicators. Perhaps the most important suggestion that these points of criticism have yielded is to consider carefully the usefulness of specific variables within the particular (national or local) educational context.

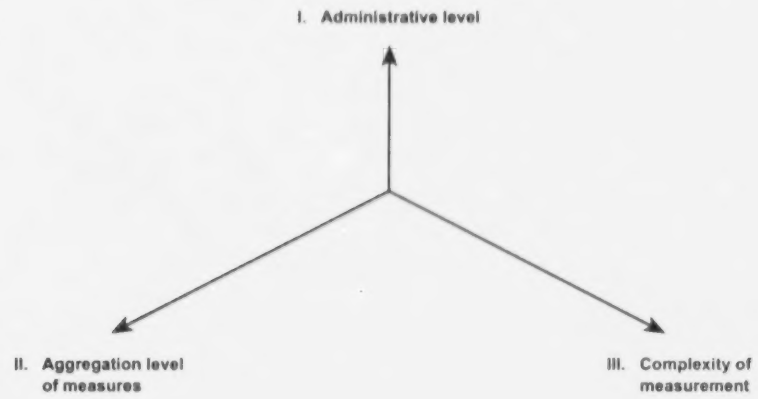
VI. MULTIPLE CONTEXT, MULTIPLE LEVEL AND MULTIPLE MEASUREMENT APPLICATIONS OF PROCESS INDICATORS

As stated in the introductory section, indicator systems can be used first in several administrative contexts: by national educational policy-makers; by officers at local level, by managers at school level and even by teachers at the classroom level.

A second distinction that has been used here is the aggregation level at which basic variables, from which indicators can be computed, are measured. In our comprehensive model of school effectiveness, we discerned measures at student, teacher/class, school and school-context level. Although we use the same levels as reference points for both distinctions, there is no compelling one-to-one correspondence between the two, since the highest administrative level may wish to use measures at any level of aggregation.

A third dimension that can be used for classifying indicators refers to the nature of measurements. Are we dealing with high- or low-inference measurements (e.g. the number of books in a school library vs. educational leadership), or are measures readily available from existing data sources? Can we use standardized quantitative scales or must we rely on relatively unstructured methods (such as participant observation or "open" interviews)? The nature of measures on which education indicators rely depends on whether indicator systems are exclusively thought of as "closed" information systems of longitudinal data or as "looser" structures that also allow for the inclusion of specific evaluative studies (such as the international comparisons of educational achievement

Figure 3. **A three-dimensional framework to classify types of indicator applications**



conducted by the International Association for the Evaluation of Educational Achievement, IEA), or in-depth studies that could be conducted by audit commissions or the inspectorate.

When we are looking at education indicator systems from the perspective of multiple administrative levels, multiple levels of aggregation of basic data and multiple types of measurement, we are clearly using a broad interpretation of the term "education indicator". The common core of all these applications still complies to the definition that was stated in the ideology section: measurement of key aspects of education systems that are of evaluative relevance.

According to the three distinctions that were discussed, we could use a three-dimensional framework to classify types of indicator applications, such as in Figure 3.

In order to determine the role of process indicators within this framework of applications, some of the most likely combinations of scale points on the three dimensions of Figure 3 will be examined.

Determining the condition of education at the national level

First, we consider decision-making at the national level supported by summary statistics at high levels of aggregation using low-inference measures and existing data bases. This is the more traditional type of application of education indicators.

Descriptive statistics on, for instance, enrolments and financial data are likely to be included in this type of application of indicators, as are statistics on educational productivity, for example, the proportion of age cohorts that pass final examinations. Could one conceive of a place for process indicators in this kind of macro-level indicator system? The aim of including process indicators would be to gain insight into possible causes for

high or low educational productivity in a particular period. The difficulty with process variables is, of course, that most of them are neither easily measurable nor readily available from existing data sources. Some proxy variables that might be considered as a basis for macro-level process indicators are:

- number of school days per year for a particular school type in a particular year;
- depending on formal regulations which require schools to register lesson "drop-out", the percentage of lessons that, for one reason or another, was not given for a particular school type, during a specific period.

These two variables could be used for calculating an indicator of net time for tuition. The policy relevance of such an indicator would depend on the variability of this indicator over time. It could also play a role in international comparisons of the functioning of national education systems.

One could also think of constructing a macro-level indicator for the evaluative potential of education systems. In this respect, it would be relevant whether or not a country has a national assessment strategy and whether or not the results of assessment are fed back to individual schools. At the national level this indicator could be of relevance because of possible changes over time, for instance when national assessment is still in a developmental phase.

A final possibility for a proxy macro-process indicator could be in the area of educational leadership by using, for instance, a measure of the total effort (time, money, enrolments) of management training courses directed at stimulating this. It should be noted, however, that such an indicator would require rather detailed content analysis of existing management training courses and therefore would probably not fit the requirement of easy measurement.

The conclusion is that process indicators have only a limited place within pure macro-level indicator systems. The possibilities are greatly enhanced if one would consider patching up these indicator systems by means of in-depth studies of process variables that could be linked to regular data streams. In-depth studies could take the form of comparative international surveys, specific research projects at the national level, or audits by expert committees or the inspectorate. In these studies, all process variables mentioned in Figure 2 could be included, be it by means of scales and questionnaire items or by means of more open check lists for observation or content analysis.

School monitoring at the district level

District-level monitoring of schools could benefit from systematic data collection as a basis for a limited set of indicators. For such management information systems, as for the previously discussed application of indicators, low-inference measures would be preferable. The most likely level of aggregation of the data would be school level, though teacher and student level data might also be used. Financial and output data (both in terms of overall productivity and student achievement) are the most important data categories for such monitoring systems. Examples of process variables that could be included are:

- whether or not schools use explicit achievement standards (to be deduced from school curricula, development plans or other official documents);

- whether or not school leaders refer to achievement records in official documents and brochures intended for the school community;
- the amount of non-teaching time that is spent on collaborative planning of curricula and lessons by the staff;
- number of students expelled from school in a particular school year;
- whether or not headteachers formally evaluate staff on a regular basis;
- the frequency of the use of achievement tests at all grade levels;
- figures on pupil absenteeism and lesson drop-out.

The above variables are assumed to be available from administrative school records or assessed by means of relatively brief encounters with headteachers. The process variables could be enlarged if research-like data were added at regular intervals.

School self-evaluation

One might say that the two applications of indicator systems that were discussed above are accountability-oriented rather than improvement-oriented. In the author's opinion, the distinction between evaluating for accountability vs. evaluating for improvement should not be drawn too sharply, since there is an important common element in both: the element of learning by means of empirical test and feedback of information to relevant actors. However, the third kind of application of education indicators that will be referred to here is most strongly associated with improvement (cf. Hopkins and Leask, 1989). Although output indicators deserve a central place in self-evaluation procedures, process variables are of particular relevance. Since organisational learning is the key motive for using self-evaluation schedules for school improvement, one will need rather detailed information on those processes that might explain disappointing results and at the same time offer suggestions for improvement. The list of process variables, based on the results of research on school effectiveness, might be used to review existing schemes and check lists for school-based review (for an overview of these check lists, see Hopkins, 1987).

One might conceive of integrating these macro-, meso- and micro-level applications of indicators. In fact, the educational data redesign project in the United States (Teauber, 1987) does exactly this. In this way detailed information is collected at the lowest unit (the classroom), and part of this information is aggregated to be used at the next level up, and so on. Although such integrated multiple-context, multiple-level and multiple-measurement indicator systems have important advantages, it is a very ambitious endeavour. One could also conceive of looser couplings of macro-meso and micro applications, for instance by exchanging instruments between the various contexts of application.

VII. CONCLUSION

The literature on school and instructional effectiveness has been used here to suggest likely candidates for process indicators. At the same time, the scope for discussing the application of process indicators was enlarged by also considering indicator applications at administrative levels below that of national governments. In Table 3 the list of variables that is proposed here as a basis for the development of process indicators is compared to other lists of process indicators that have been proposed.

As appears in Table 3, there is considerable agreement between the proposal that is made here and those by Oakes, Benveniste and Teauber. The proposals by UNESCO and Windham are stated in more general terms. The one factor that is included in all of the proposals is instructional time. The categories of variables that distinguish the proposal that is made here from the others are: achievement stimulants from the school environment, structured teaching and the evaluative potential of the school.

Process indicators are a somewhat difficult category among other, more established types of indicators. The main difficulty is that process indicators usually require rather complicated procedures of data collection and measurement and thus do not fit the requirement of readily available data for computing indicators. Yet, there is an important motive to try and include process information in indicator systems. Like data on resources and inputs, process data provide background material that is helpful in making sense out of mere performance data on the functioning of education systems.

Table 3. **Comparison of sets of process indicators**

Scheerens (1989)

Achievement stimulants
Achievement-oriented policy
Educational leadership
Teachers' co-operative planning
Quality of curriculum
Evaluative potential
Orderly climate
Time on task
Structured teaching
Opportunity to learn
High expectations
Monitoring progress
Reinforcement

Windham (1988)

Instructional organisation
Alternative technologies
Use of teacher and student time

Benveniste (1987)

Teacher time (teaching/non-teaching)
Student learning time:
- course enrolment
- turnover rates
- pupil/teacher ratios
- school day activities
- length of school year
- out-of-school learning time

Order and consistency:
- truancy, absenteeism, vandalism, disruptions
- student turnover
- student co-operative behaviour

Teauber (1987)

Instructional leadership
Curriculum
Type of instruction (whole class, small group, etc.)
Time on task
School climate
Influence of peer group

Oakes (1987)

Access to knowledge (e.g. instructional time)
Press for achievement (e.g. graduation requirements)
Professional conditions for teaching (e.g. time spent on collaborative planning)

UNESCO (1976)

Allocation of resources
Retention and progression rates
Teacher hours per pupil per year
Cost and management

BEST COPY AVAILABLE

BIBLIOGRAPHY

Achilles, C.M. and M.N. Lintz (1986), "Evaluation of an effective schools intervention", American Educational Research Association (AERA) paper, San Francisco.

Averch, H.A. *et al.* (1974), *How Effective is Schooling?: A Critical Review of Research*, A Rand educational policy study, Englewood Cliffs: Educational Technology Publications.

Benveniste, G. (1987), "The design of school accountability systems", paper presented at the Washington conference on education indicators.

Bosker, R.J. and J. Scheerens (1989), "Criterion choice, effect size and stability. Three fundamental problems in research on school effectiveness", in B.P.M. Creemers and B. Reynolds (eds.), *School Effectiveness and School Improvement*, Lisse: Swets and Zeitlinger.

Bossert, S.T. (1988), "School effects", in N.J. Boyan (ed.), *Handbook of Research on Educational Administration*, New York: Longman Inc.

Breton, A. and R. Wintrobe (1982), *The Logic of Bureaucratic Conduct*, Cambridge: Cambridge University Press.

Brookover, W.B., C. Beady and P. Flood *et al.* (1979), *School Social Systems and Student Achievement – Schools can Make a Difference*, New York: Praeger Publishers.

Brophy, J. and Th.L. Good (1986), "Teacher behavior and student achievement", in M.C. Wittrock (ed.), *Handbook of Research on Teaching*, New York: McMillan Inc. pp. 328-375).

Cameron, K.S. and D.A. Whetten (eds.) (1983), *Organisational Effectiveness. A Comparison of Multiple Models*, New York: Academic Press.

Coleman, J.S. *et al.* (1966), *Equality of Educational Opportunity*, Washington D.C.: US Government Printing Office.

Collins, A. and A. Stevens (1982), "Goals and strategies of inquiry teachers", *Advances in Instructional Psychology, Vol. II*, R. Glaser (ed.), Hillsdale, N.J.: Lawrence Erlbaum Associates.

Doyle, W. (1985), "Effective secondary classroom practices", in M.J. Kyle (ed.), *Reaching for Excellence. An Effective Schools Sourcebook*, Washington: US Government Printing Office.

Faerman, S.R. and R.E. Quinn (1985), "Effectiveness: the perspective from organisation theory", *The Review of Higher Education*, 9, 83-100.

Firestone, W.A. and R.E. Herriott (1982), "Prescription for effective elementary schools don't fit secondary schools", *Educational Leadership*, 40, 51-53.

Fraser, B.J., H.J. Walberg, W.W. welch and J.A. Hattie (1987), *Syntheses of Educational Productivity Research*, Special Issue of the *International Journal of Educational Research*, Vol. 11, No. 2.

Friedkin, N.E. and J. Necochea (1988), "School system size and performance: a contingency perspective", *Educational Evaluation and Policy Analysis*, 10. 237-249.

Glasman, N.S. and J. Biniaminov (1981), "Input-output analysis of schools", *Review of Educational Research*, 51, 509-539.

Hanushek, E.A. (1979), "Conceptual and empirical issues in the estimation of educational production functions", *Journal of Human Resources*, 14, 351-388.

Hanushek, E.A. (1986), "The economics of schooling: production and efficiency in public schools", *Journal of Economic Literature*, 24, 1141-1177.

Hauser, R.M., W.H. Sewell and D.F. Alwin (1976), "High school effects on achievement", in W.H. Sewell, R.M. Hauser and D.L. Featherman (eds.), *Schooling and Achievement in American Society*, New York: Academic Press.

Hopkins, D. (1987), *Doing School Based Review*, Leuven, Belgium: ACCO.

Hopkins, D. and M. Leask (1989), "Performance indicators and school development", *School Organisation*, Vol. 9, No. 1: 3-20.

Jencks, C. *et al.* (1972), *Inequality*, New York: Basic Books.

Jencks, C. *et al.* (1979), *Who Gets Ahead? The Determinants of Economic Success in America*, New York: Basic Books.

Kulik, C.L.C. and J.A. Kulik (1982), "Effects of ability grouping on secondary school students: a meta-analysis of research findings", *American Educational Research Journal*, 19, 415-428.

Kyle, M.J. (ed.) (1985), *Reaching for Excellence. An Effective Schools Sourcebook*, Washington D.C.: US Government Printing Office.

Levin, H.H. (1988), "Cost-effectiveness and educational policy", *Educational Evaluation and Policy Analysis*, 10, 51-69.

Lezotte, L.W., R. Edmonds and G. Ratner (1974), "A final report: remedy for school failure to equitably deliver school basic skills", Michigan State University.

McCormack-Larkin, M. (1985), "Ingredients of a successful school effectiveness project", *Educational Leadership*, March, 31-37.

Miller, S.K., S.R. Cohen and K.A. Sayre (1985), "Significant achievement games using the effective schools model", *Educational Leadership*, March, 28-43.

Mintzberg, H. (1979), *The Structuring of Organisations*, Englewood Cliffs: Prentice Hall.

Mortimore, P., P. Sammons, L. Stoll, D. Lewis and R. Ecob (1988), *School Matters: The Junior Years*, Somerset: Open Books.

Mosteller, F. and D.D. Moynihan (eds.) (1972), *On Equality of Educational Opportunity*, New York: Random House.

Niskanen, W.A. (1971), *Bureaucracy and Representative Government*, Chicago: Aldine-Atherton.

Oakes, J. (1987), "Conceptual and measurement issues in the construction of school quality indicators", paper presented at the AERA-annual meeting, Washington D.C.: American Educational Research Association.

Purkey, S.C. and M.S. Smith (1983), "Effective schools: a review", *The Elementary School Journal*, 83, 427-452.

Ralph, J.H. and J. Fennessey (1983), "Science or reform: some questions about the effective schools model", *Phi Delta Kappan*, 689-695.

Rutter, M. (1983), "School effects on pupil progress: research findings and policy implications", *Child Development*, 54, 1-29.

Scheerens, J. (1989), *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*, Den Haag: Stichting voor Onderzoek van het Onderwijs.

Scheerens, J. and W.G.R. Stoel (1988), "Theory formation on school effectiveness", paper presented at the annual meeting of the American Educational Research Association (AERA), New Orleans.

Scheerens, J., W.G.R. Stoel, C.J.A.J. Vermeulen and W.J. Pelgrum (1988), *De haalbaarheid van een indicatorenstelsel voor het basis- en voortgezet onderwijs*, Enschede: Universiteit Twente.

Scheerens, J., H.C.R. Nanninga and W.J. Pelgrum (1989), "Generaliability of instructional and school effectiveness indicators across nations", in B.P.M. Creemers and J. Scheerens (eds.), *Developments in School Effectiveness Research*, Special Issue of the *International Journal of Educational Research*, Vol. 13.

Stern, J.D. (1986), "The educational indicators project at the US Department of Education", Center for Statistics, US Department of Education.

Teauber, R.C. (ed.) (1987), *Education Data System Redesign*, Special Issue of the *International Journal of Educational Research*, Vol. 11, No. 4: 391-511.

Teddlie, C., S. Stringfield and R. Wimpelberg (1987), "Contextual differences in effective schooling in Louisiana", Washington: American Educational Research Association (AERA).

Thompson, J.D. (1967), *Organisations in Action*, New York: McGraw Hill.

Thorndike, R.L. (1973), *Reading Comprehension Education in Fifteen Countries*, Stockholm: Almqvist and Wiksell.

UNESCO (1976), "Towards a system of educational indicators", Division of Educational Policy and Planning, Paris.

Walberg, H.J. (1984), "Improving the productivity of American schools", *Educational Leadership*, 41, 19-27.

Weber, G. (1971), "Inner-city children can be taught to read: four successful schools", Washington D.C.: Council for Basic Education.

Windham, D.M. (1988), *Effectiveness Indicators in the Economic Analysis of Educational Activities*, Special Issue of the *International Journal of Educational Research*, Vol. 12, No. 6.

Chapter 4

DO COMMON VALUES PRODUCE COMMON INDICATORS?

by

Alan Ruby
Department of Employment, Education and Training
Canberra, Australia

I. INTRODUCTION

In the last five years there has been considerable interest both nationally and internationally in devising education indicator systems. The underlying assumption of models of education indicator systems is that nations, provinces, school systems and educational institutions "share" some concerns and goals, so they can be meaningfully compared. Without these shared values the indicators lack policy relevance and technical vigour because they have no clear frame of reference.

The common concerns that have influenced work on internationally comparable indicators are readily identifiable. Indeed they dominated the agendas of the Washington and Poitiers Conferences. Concerns such as *i)* demands for more public and political accountability in education; *ii)* the re-emergence of the links between economic competitiveness and quality of educational services as a dominant policy value; and *iii)* the press for greater equity in educational outcomes are common phenomena in all industrialised nations (Ruby, Millan and Howe, 1989).

These concerns and other shared social and political phenomena have shaped the demand for better and more accessible information. Prior to committing to an investment in indicator development (Ruby, 1989) at a conceptual and theoretical level, it is also often assumed that education systems are sufficiently similar or share enough values to be readily or meaningfully compared. Van Herpen's survey of the various models for indicator systems (Chapter 2) developed in the last 15 years shows that structural and philosophical similarities are assumed either explicitly or implicitly in nearly all models. He concludes that the variation in models has largely been confined to differences in the role ascribed to environmental factors in education.

Even the most recent theoretical work (for example Nuttall in Chapter 1) makes the assumption that schools, the educational process itself and school or education systems are essentially comparable. Nuttall's very definition of an indicator includes the idea of monitoring the "enduring features of the system" – implying that there are fixed or common elements.

In practical terms it is easy to make, and support, assumptions about common values, structures and processes as a basis for the development of indicator systems. At the macro level the very existence of formal school systems, the involvement of government in education, the notion of compulsory schooling and the enduring delivery model of a teacher being responsible for a group or class of students, support the assumption of shared values. The untested questions are: Can these values be identified, defined and transformed into indicators in a way in which all, or many, countries or commentators will agree that they are appropriate representations of the situation in their country or schools? Can we move from the broad general values to specific operational measures?

These questions can be examined by critically appraising some of the work on indicator development undertaken during Phase I of the OECD/CERI Project. While the first phase of this work has been largely successful, the developmental work did reveal that there were significant problems to be overcome in deriving agreed definitions of indicators. There were conceptual, technical and practical problems. All three sets of problems are well illustrated in an example from work on student flows (Ruby *et al.*, 1989).

II. TRANSLATING VALUES INTO MEASURES

In seeking to derive indicators of student flows we tried to operationalise the common goal of regular attendance at school. Perhaps naively we had assumed that regular attendance at school was a common and important policy goal in all OECD countries. We rationalised our choice saying that common sense predicted that if a child went to school every day, the chances were higher that he or she would learn more than the student who was often absent.

We also knew from some research that children who had poor attendance in the early school years, often described as compulsory years, were most likely to "drop out" later on, frustrating the goal of greater participation. Indeed, it is still safe to assume that OECD countries do value regular attendance at school.

The difficulty comes when, having identified a common value, we proposed to measure system performance in regard to that value. Very few countries were able to provide data that measured the phenomenon. Why was this so?

The reasons offered by participating countries are summarised in the explanatory table. Some countries offered more than one reason, and some commented and also provided data. Analysing these reasons, it is possible to discern different sets of explanations.

One set of explanations is that school attendance is not a policy or political problem (Switzerland and Japan). It may be that the culture of some countries places much greater significance on regular school attendance or supports school attendance through other social institutions and practices to a degree that limits absenteeism to such an extent that it is not a policy problem. Conversely, some countries may sanction non-attendance – particularly for the Saturday half-day, common in some OECD countries.

Another set of explanations is that school attendance is simply a low priority for international comparison. There are other factors which affect total time spent in learning – such as teacher absenteeism and length of school year and day (Canada, the Netherlands).

Explanations about non-availability of attendance data

1. "Absenteeism is no issue... neither public opinion nor the education administrators are concerned about absenteeism, and it is totally missing from public or administrative debate" (Switzerland).
2. "... our government does not have almost any political interest in the attendance rate" (Japan).
3. "...not attending school is forbidden by law but the law is not enforced (Belgium).
4. "Not a priority in the perspective of international comparisons. However, the formal duration of the school year, as well as the extent of the weekly schedule, remain valid process indicators" (Canada).
5. "Research discovered that 15 to 20 per cent of all lessons in secondary education are not given because of sickness (teacher), examinations, or staff meetings. This seems to be a far more serious problem than the absenteeism of pupils..." (the Netherlands).
6. "If it exists, then only on a marginal base" (Switzerland).
7. "It can be concluded from these data (on students absent for more than 50 days) that the attendance rates... in general are fairly high" (Japan).
8. "The attendance rates for these grades (compulsory) only depend on illness. Otherwise it would be 100 per cent. Therefore... doesn't apply" (Federal Republic of Germany).
9. "The attention is mainly given to long-term absenteeism... this is directed to monitoring problem-pupils: potential drop-outs" (the Netherlands).
10. "Information is only available at the school level" (Belgium).
11. "There is some information but the quality differs between the municipalities" (the Netherlands).
12. "Our data vary considerably from state to state, so we have difficulty in producing good national estimates" (United States).
13. "It is impossible to obtain attendance rates at this period of the school year" (Belgium).

A further set of explanations focuses on the associated problem of drop-outs or chronic absenteeism (the Netherlands). The argument is to focus on those most at risk.

Some countries do see attendance as a policy problem but do not hold data at the national level. This is understandable. The school or district level is where action to limit/prevent/follow-up absenteeism takes place. The data are held at the level where improvement can be initiated. This can give a slightly mixed message – that attendance is not a national concern or a high-level social value.

These explanations are not mutually exclusive. Some are obviously related and some countries offered more than one of these explanations as to why they did not hold these data. It may also be that some countries see absenteeism as an intractable problem and, hence, too politically difficult for regular data collection on an internationally comparable basis.

This diversity of views suggests prior questions: Why should we collect data on attendance, or its corollary, absence? What policy issues can be informed by analysis of these data? What research is there to argue for their inclusion in an indicator system?

Attendance data can be used to guide the distribution of resources, particularly marginal resources. Some school systems have allocated personnel to improve attendance-monitoring procedures.

There is a small body of research that shows that absenteeism is correlated with student drop-out both on an individual and school basis. A recent Canadian study

(Lawton *et al.*, 1988) also suggested that students themselves link attendance with performance. The same study recommended provincial analysis of drop-out data and greater use of attendance-monitoring systems.

The related issue of partial absenteeism, missing classes or part of the school day, is also a well researched phenomenon. It too is associated with school performance. Again, the common sense notion is that students are more likely to learn something if they have been taught it. Success in learning, particularly success in terms defined by the school, is also closely linked with school completion or graduation.

In these simple terms, absenteeism links with the broader sense of participation in schooling which we adopted for our work on indicators. It too is a shared value, common to many school systems.

III. THEN WHY IS IT SO DIFFICULT?

What is it then that makes it difficult to move from the agreed goals to agreed indicators? What factors impede the translation of these principles into an effective indicator system? There are three obvious factors.

First, there are fundamental differences in the cultural significance and the social interpretations placed on the agreed values. The degree of individual adherence to these values also varies markedly between countries. More heterogeneous nations may have greater variation in attitudes towards overarching social goals and may tolerate greater diversity in responses to those goals. These differences influence the relative incidence of social behaviours such as absenteeism. They also influence the readiness of governments to acknowledge the existence of such behaviours – the willingness to identify and quantify a policy problem.

Second, different philosophies in regard to policy and action also affect the implementation of models of indicator systems. Some countries believe that the responsibility for policy and action on such issues as attendance rests at the site level and that data to help discharge that responsibility should not or need not go past that level.

Other countries see attendance as a state or system level responsibility and establish data collection strategies to monitor school and district attendance rates. The application of these philosophies is not uniform within countries but varies with the value question concerned. It is also affected by the governance structure of education within countries. Highly decentralised systems with high levels of local policy determination tend to have few data on attendance. The corollary does not necessarily apply – cultural issues can mediate the demand for national data collection.

The third factor is slippage or dissonance between the values and interests shaping national or system policies and the priorities reflected in data systems. This can be a product of organisational arrangements, of the absence of political will, shortcomings in technical capacity and different time frames. This last is most interesting: the value of data systems comes in part from their historical basis – the capacity to show change over time. The cost of data systems also dictates that they continue largely unchanged for some years. Policy problems often emerge quickly and almost always demand immediate answers. This produces a tension between policy-makers and data managers. Indicators are one strategy for relieving this tension and the models are the organising framework for this to happen.

IV. CONCLUSION

The unresolved question is: What should guide our choice of indicators? Do we select the common policy problems or do we focus on the enduring values that underpin our education systems? This assumes that there are either sufficient common policy issues or shared values to underpin a system.

It would be foolish to devise an indicator system without addressing both elements. The aim should be to balance the two sets of concerns and do so without overloading an indicator system with too many data. To constrain the scope of systems we need to put the large and fixed values such as attendance under periodic review – say, every five to ten years – rather than annual. Similarly, we need to make more use of and promote the acceptance of sample data. These techniques will help us resolve the competing demands but, most of all, we need to establish protocols for selecting and interpreting indicators from the set. This is the next step in the conceptual development of indicators and an essential precursor to realising the benefit of internationally comparable indicators.

Acknowledgements

The author wishes to thank Brian Davies, of the Australian Indicators Project Team, and Isabelle Delfau, of the CERI Secretariat, for their assistance in the preparation of this chapter.

BIBLIOGRAPHY

Lawton, S.R., *et al.* (1988), *Student Retention and Transition in Ontario High Schools: Policies, Practices and Prospects,* Ministry of Education, Ontario.

Ruby, A. (1989), "Why Invest in Performance Indicator Development?", in Wyatt, T. (ed.), *Reporting Educational Progress: Performance Indicators in Education,* Bulletin 11, Sydney: Conference of Directors-General.

Ruby, A., M. Millan, and V. Howe, (1989), "Indicators and Equity: The Neglected Part of the Debate", in Wyatt, T. (ed.), *Reporting Educational Progress: Performance Indicators in Education,* Bulletin 13, Monograph 7, Sydney: Conference of Directors-General.

Ruby, A., *et al.*, Network 1 (1989), *Enrolments, Career Paths and School Leavers at Various Stages,* Network Report, Sydney, Conference of Directors-General, August.

Chapter 5

STATISTICAL AND INDICATORS WORK IN THE OECD DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY

by

Gunnar Westholm
OECD Secretariat
Directorate for Science, Technology and Industry

I. INTRODUCTION

Science and Technology (S&T) indicators and/or Research and Development (R&D) statistics are by now widely accepted and utilised tools at all levels in national and international science policy systems. The OECD has, over time, become a leader in the field of such indicators.

This chapter summarises the efforts made to develop internationally comparable statistics and indicators of the "inputs" of human and financial resources devoted to S&T/R&D and of the "results/outputs" of R&D in the framework of the Organisation's programme on scientific, technological and industrial indicators.

Specific emphasis will be given to the "philosophy" or the rationale for collecting such data and developing such indicators, and to the importance of establishing and maintaining contacts with users and producers of data at the national level and with the OECD policy committees. The various short-, medium- and long-term stages in the development of indicators will be discussed in terms of methodological and practical data collection, processing and dissemination activities.

II. SUMMARY HISTORY OF S&T INDICATORS

Background

Science and technology indicators have a number of origins and predecessors. Already in the inter-war period, following the lead of the German chemical industry and the American electrical industries, professional R&D departments were becoming a regular feature of the industrial landscape and the first efforts to measure this phenomenon were made by industrial federations or by independent academics.

The post-war revival in the theory of economic growth did not pay any attention to R&D until the mid-1950s with the identification of the "residual factor" in studies which attributed an important role to "better ways of making things". At about this period the concept of "Science Policy" emerged in OECD Member countries, accelerated by the shock of Sputnik in 1957.

Interest concentrated on two types of indicators, the quality of the labour force and resources devoted to R&D. The OECD entered the field of R&D statistics in 1961 with a pilot study followed by a draft Manual and the first of its biennial international surveys. This period of intense interest in R&D, mainly in an economic context, continued until the late 1960s.

In the 1970s, R&D reverted to being a matter for specialists as, on the one hand, policy-makers became more interested in social, environmental and then energy questions and, on the other, economists discovered that global econometric comparisons between R&D statistics and economic data gave very inconclusive results. Experimental efforts were made to identify possible measures of the results and outcomes of R&D, with the United States leading the field.

In 1980 the OECD organised a major conference to bring together the main players working on these new "science and technology indicators" such as patents, the technological balance of payments, bibliometrics and innovation studies. During the 1980s, the OECD has been working towards establishing norms and surveys for these new indicators. At the same time, there has been a revival of interest by economists, this time mainly in the context of industrial adjustment and competitiveness. Stress is now laid on measurement at the level of individual industries and on taking into consideration their total technological resources including not only R&D but also the skills of the workforce, technology transferred in the form of licences, patents, materials and equipment, and all the costs of successful innovation.

There are thus multiple opportunities for developing new, policy relevant S&T indicators. The OECD is currently undertaking an Organisation-wide exercise entitled the Technology-Economy Programme (TEP) for which eight key areas of analysis have been identified:

- science, technology and innovation;
- diffusion and techno-economic networks;
- technology and investment;
- technology and human resources;
- technology and productivity;
- technology and competitiveness;
- technology and environment;
- technology and development.

On the occasion of TEP-related seminars and colloquia in national capitals, issues of relevance to S&T indicators were also at stake. The consequences of the TEP for the development of indicators were discussed at a special conference in Paris in July 1990.

S&T indicators in the OECD context

The OECD does not have a "central statistical office". Data are collected in the various mission-oriented Directorates to meet the needs of their respective programmes

of work and policy committees. In the mid-1980s the former Science and Technology Indicators Unit was upgraded to the Scientific, Technological and Industrial Indicators Division (STIID) which currently has a staff of twenty-three.

The major responsibility of the STIID is to carry out the Joint Statistical Programme of Work for three OECD Committees: the Committee for Science and Technology Policy, the Industry Committee and the Committee for Information, Computer and Communications Policy. The work of the STIID includes carrying out major statistical projects, developing indicators to meet policy needs, providing statistical expertise to other divisions and providing the secretariat to the various statistical subgroups of the three policy committees it is serving. The STIID also facilitates the international exchange of ideas on the development of international standards.

In order to establish a balance between the needs for statistical work of the three policy committees mentioned and the limited resources available, an Informal Group on the Joint Statistical Programme of Work of representatives from each committee has been set up to decide on the short-term co-ordination of STIID work.

III. THE PHILOSOPHY BEHIND THE SELECTION AND DEVELOPMENT OF INDICATORS

General

Collecting statistics and developing indicators is not a purpose *per se*. Indicators are meant to answer questions and if necessary to give "early warnings". It is clear that the need for indicators differs considerably between various categories of users and that most often a single indicator is not sufficient to give relevant information. This is true both for "input" and for "output" indicators. It is frequently necessary to apply a combination of various indicators to have a reasonably accurate response to the questions put forward.

Should an agency such as the OECD collect the basic statistics from Member countries or go directly for the indicators? It has always been the practice, in the case of S&T data, to collect the statistics and calculate the indicators at the OECD. For example, perhaps the best known single R&D indicator is R&D as a percentage of Gross Domestic Product (GDP). The STIID collects standardized data on Gross Domestic Expenditure on R&D (GERD) from national specialists and calculates the percentage using standardized GDP figures provided by the OECD Economic Statistics and National Accounts Division, hence guaranteeing maximum international comparability.

One important aspect of the Technology-Economy Programme (TEP) indicators project mentioned above is the definition of a theoretical framework for the design and selection of new science and technology indicators. This section draws on this to present a number of definitions, categories and criteria which are rather abstract and thus applicable to other types of indicators.

Definitions

Science and Technology Indicators can be defined as "series of data designed to answer questions about the Science and Technology System (STS), its internal structure,

its relation with the economy, the environment and society, and the degree to which it is meeting the goals of those who manage it, work within it or are otherwise affected by its impacts".

Statistics are the basic material (atoms) from which indicators (molecules) are constructed. The questions which indicators should answer concern those aspects of wider issues which can be examined using quantitative techniques.

Categories of statistics and indicators

Both statistics and indicators can be divided into two categories:

- *Standard indicators* are based on structured data bases reflecting accepted opinions about the pattern and processes of the STS and the economy, and are designed to answer the most common questions.
- *Specific indicators* are designed to answer specific questions about the STS and usually require special data collection exercises to obtain the basic statistics. These questions vary over time and between countries because of changing or differing policy objectives, industrial specialisation, etc. Sometimes the same "special needs" occur in a number of countries over several years (e.g. interest in information technology indicators) and enter the OECD ambit.

Criteria for the development of indicators

The following criteria are of primary interest in the development of internationally comparable indicators:

i) Importance

International demand: The indicators must address issues of interest to the majority of national capitals, whether they answer policy or analytical questions.

International value-added: OECD indicators must permit analysis and policy conclusions over and above what can be done in national capitals.

Sustained demand: As a rule international indicators should not be collected for purely temporary phenomena.

ii) Quality

Theoretical soundness: The indicator's operational value, i.e. in particular its causal validity and conceptual relevance, must be acknowledged by specialists in the field.

Reliability: The indicator serves as a basis for a policy decision which may have substantial socio-economic repercussions, and must accordingly be all the more reliable. In this respect, one should not confuse long-standing use with reliability but must check the accuracy of concepts or methods which may have become obsolete, inappropriate or irrelevant.

International comparability: Assessing the country's structural assets or its performance level means that they have to be compared with those of coun-

tries of the same size or with those with which comparisons are particularly desirable (trading partners, etc.).

Dynamism: Indicators must be comparable over time.

iii) Appropriateness

Operationality: Here the reference is to the appropriate actor at the appropriate level. The indicator must thus allow an effective policy action to be taken at least at the national level but also at other relevant levels.

Directness: The indicator must be as direct as possible so that policy action or analytical exercise itself has the greatest possible chance of directly achieving its objective. The indicator therefore must not be too synthetic.

Adaptation: The indicators must be adapted:

- To the user: if the indicator is to enable governments to take a policy decision, it has to be geared to the decision-making body concerned (international, national, regional or other organisation);
- To the objective: and accordingly be practicable at macro-economic, meso-economic, or micro-economic level;
- To the development stage in innovation.

iv) Availability

Timeliness: Timeliness will differ according to the type of action to be considered: availability at short notice for short-term action, etc.

Realisability: There must be a reasonable possibility of collecting the basic statistics and cracking the main methodological problems.

Resource-efficiency: The cost of data collection by national authorities, its reporting to OECD, its assembly, verification and methodological work must be taken into consideration as must the cost to the agency of buying data from commercial sources.

IV. MAINTAINING CONTACTS WITH USERS OF INDICATORS AND PRODUCERS OF DATA

It is important to identify the main categories of present or potential users of indicators and also necessary to recognise the role of current and potential suppliers of data to attain an optimal adaptation of the indicators to the needs of the former, at a reasonable cost (in time and money resources) to the latter.

The following paragraphs will discuss some of the considerations defined within the framework of the OECD Technology-Economy Programme (TEP) with respect to the relationships between the Secretariat and users and producers of its statistics, indicators and information. The section closes with a more practical summary of lessons learned from work with S&T indicators.

Users of S&T indicators and their characteristics

Four potential groups of users of S&T indicators have been identified:

- policy-makers;
- analysts;
- persons working within the Science and Technology System (STS); and,
- the general public.

All four groups are usually interested in the same broad issues, for example the effect of technological change on employment or the environment, but they ask somewhat different questions, participate to varying degrees in the preparation of these indicators and may interpret them in different ways.

i) Policy-makers

Work on S&T indicators is mainly intended to serve policy-makers. Ministers themselves usually require information in a readily usable form such as a graph and thus are not considered to be direct consumers of S&T indicators.

Most of the S&T statistics and indicators collected and issued by the OECD are designed for use by the staff who prepare decisions for ministers. For this, they need three types of indicators:

- A "scoreboard" of standard indicators which allows them to compare their nation's scientific, technological and industrial performance and competitiveness over time and with other countries;
- Specific indicators to inform them about special problem areas;
- Specific indicators to measure the degree to which the STS is reaching the goals towards which the policy-maker is trying to steer it, either concerning that part of the STS which is government-managed or industrial science, technology and industrial activities.

These users are represented on OECD Committees such as the Committee for Scientific and Technological Policy, the Industry Committee or the Committee for Information, Computer and Communications Policy.

ii) Analysts

Analysts are usually academics and/or members of government policy teams. They are generally in touch with the OECD at specialised workshops or as consultants. They usually prefer to build their own indicators from a stock of statistics. They pioneer work at the cutting edge of S&T indicator development and often have to invent new methods and make imaginative use of new sources and types of statistics in order to advance. However, most analytical work in policy offices or at the OECD involves redeploying existing or improved statistics to devise indicators to test current hypotheses on a larger number of cases, industries or countries.

Once the analysts' hypotheses have been thoroughly tested and are generally accepted, they become part of the conceptual framework for standard indicators work, underlying, for example, the "scoreboard".

iii) Persons working within the STS

This category covers all those persons actually carrying out R&D, technological and related industrial activities.

There are some very important S&T policy-makers in the private sector, particularly at the international level. For example, as is well known, IBM commits as much money to R&D as the Italian government. The TEP has offered a welcome opportunity to find out what kind of S&T indicators such industrialists would find interesting. The STIID already receives a large number of requests for data from industry, usually concerning individual industries or even for products, for which the Division's data bases do not include sufficiently detailed series.

Scientists and engineers employed in the STS have a legitimate interest both in overall trends and in what is happening in their own field. Some are looking for information about their specific areas in order to identify recent trends or to situate their projects in a slightly wider context. (Such users generally need data for individual fields of science, technology or industry, often in more detail than the OECD system can supply). Others are seeking indicators in order to justify their projects and their need for more funds.

iv) The general public

Public attitudes towards science and technology are formed by a number of factors including articles in the press, television and radio programmes, etc. Supply of accurate and attractively presented S&T indicators to the media could contribute to a better understanding of the STS by the public at large. Conversely, data on public attitudes to science and technology can assist with planning, notably for S&T education and, at the other end of the scale, for marketing new science-based goods and services.

The producers of S&T data

i) Government STI statistical services

These are the bodies whose primary task is to collect and issue government scientific, technological or industrial statistics. Most of these are part of central statistical offices though some are special units in S&T ministries (notably those preparing "Science budgets"), in industry ministries or in research councils. Some, though not all, also design and use standard indicators.

ii) Other public bodies

These are bodies which collect and issue data intended for other purposes but which are also used to construct S&T indicators. Notable examples are patent offices and central banks (for Technology Balance of Payments – TBP – data).

iii) Commercial sources

These are companies or consultants who collect and issue statistics or who produce indicators as a commercial service. Such commercial suppliers dominate the market in the case of detailed and up-to-date statistics on information technologies (IT) products

and at the other end of the spectrum, regarding bibliometrics data on publications and citations.

V. STAGES OF WORK TO DEVELOP NEW STATISTICS AND INDICATORS

General

As new indicators are developed they go through a number of stages, each with different characteristics and each requiring a different mix of resources both in the participating countries and at the OECD. Of course, each project will follow a slightly different trajectory but for planning purposes it is possible to identify three main stages: "experimental", "developmental" and "operational".

The standard OECD methodology for measuring the resources devoted to R&D in Member countries is that of the *The Measurement of Scientific and Technical Activities. Proposed Standard Practice for Surveys of Research and Experimental Development*, best known as the "Frascati Manual" (current 4th edition, 1980). The fifth revision meeting took place in Rome in the Autumn of 1991. The history of this Manual is a good example of the three stages just mentioned, with the experimental phase starting in the early 1960s, followed by the first international survey covering 1963/1964 and then, from 1967 onwards, regular biennial OECD surveys with increasing participation of Member countries, in combination with a series of methodological revisions based on the resulting practical experience both at the OECD and in the Member countries. (The "history" of the Frascati Manual is summarised in Annex 1).

The experimental stage

During the experimental stage, the need for a new type or set of indicators is identified and the first attempts are made to collect them in a few countries, often by academics for use in their own analytical work. Both agencies in national capitals and the OECD statistical services need to have their antennae tuned to pick up these new possibilities.

In the case of S&T indicators, the OECD was able to play a useful role by organising workshops to which all those with relevant experience were invited to come and exchange their views and experiences. At the very beginning of the process, the participants are mainly academics and the output of the workshops is a mass of rather theoretical papers, plus an invisible college of specialists and informed official experts who attend as observers. Towards the end of this period, sufficient data and experience have been built up to undertake a pilot study and draft some initial guidelines for future surveys. At this stage, before a major project can be inserted into the OECD programme of work, the lead country approach can be extremely useful. For example, the Nordic countries are currently very active in developing innovation indicators. Subsequently a consultant, often an academic, is retained to draft the first guidelines.

Most of the experimental work on R&D indicators took place in the early 1960s though some additional series went through the stage later. However, there has been a mass of experimental work on measures of the output and impact of R&D since the late 1970s with a serious problem of deciding which should be carried forward from the

experimental to the development stage. Two typical cases are the measurement of stocks of scientific and technical personnel on which an initial workshop was held in 1981, and bibliometrics which has been discussed at regular intervals at workshops since the late 1970s.

The development stage

This stage generally begins with the organisation of the first OECD survey based on international guidelines. This involves work in national capitals to fill in the new questionnaire and at the OECD to set up the initial data base and programme to check the responses for obvious arithmetic errors. However, the key requirement is for a very careful assessment of the quality and comparability of the responses by qualified and experienced staff at the OECD. This is unfortunately a very time-consuming but absolutely necessary task in order to build up, from the very beginning, a consistent data base. Problems may arise when respondents have not understood the guidelines, because the guidelines are not clear or because responses are affected by institutional or cultural differences between countries. The first OECD R&D survey concerned the year 1963 or 1964. For most of the seventeen participating countries this was a new experience and it was not surprising that the quality and comparability of their responses varied considerably.

Where there are important differences between countries, close contact with national survey agencies is essential. However, the resource limitations of both the OECD and Member countries often restrict such contact.

The operational stage

The operational stage can be said to be reached, in the OECD Secretariat and in Member countries, when surveys are regularly undertaken using operational questionnaires and when routines have been established for sending out and receiving them for the processing, verification and dissemination of the statistics and when quality control of data for comparisons both in space and in time is carried out. Here, only marginal improvements are possible to increase the efficiency.

This stage requires a major initial investment in computer programming in order to set up detailed data bases and the related sets of verification and treatment programmes based on the experience of the preceding stage, plus the programmes to produce any associated regular publications, diskettes, tapes, etc. Only then can the amount of professional time be reduced and an increasing role for assistant staff becomes appropriate and, indeed, essential. This stage has now been reached for OECD R&D statistics.

VI. FURTHER DETAILS ON STIID ACTIVITIES

General

At each of the above stages it is essential to move forward on the three broad fronts of methodology, data collection and diffusion, bearing in mind that the borderlines

between the three tasks are difficult to define. They are all essential and it may sometimes be hard to find a proper balance between them.

Methodology

OECD/S&T manuals

The "Frascati Manual" is, so far, the best known of the methods developed by the OECD for international comparisons of science resources. Its basic definitions of R&D have been adopted also by other international organisations, such as UNESCO (which co-operated closely with the OECD, notably during the first years of the development of the Manual), the European Communities and the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) in their former science-policy group "Nordforsk".

The Manual and its earlier versions have been translated – as well as appearing in the official OECD English and French versions – into Dutch, German, Italian, Japanese, Portuguese (in Brazil), Serbo-Croat, Spanish, Turkish and, apparently, recently also into Chinese. Several non-Member countries (recently including Eastern European countries) have contacted the OECD Secretariat to learn more about the guidelines and at least one such country (Singapore) makes direct reference to Frascati practice when undertaking its own national R&D survey.

Methodological work does not necessarily involve collection of new data series but may just as well centre on developing methods of exploiting already available data for new indicator purposes. This is the case of the work currently carried out in the STIID in close contact with the Member countries to prepare manuals, similar to the "Frascati Manual", for the use of Technology Balance of Payments statistics (the "Proposed Standard Method of Compiling and Interpreting Technology Balance of Payments Data – TBP Manual 1990" was issued in 1990), of patents data and of "high", "medium" and "low" technology trade and production statistics as science and technology "output" indicators.

Work is well advanced on the preparation of a manual for the collection of ICC (information, computer and communications) statistics. Also being developed is an OECD methodology for the collection of data on innovation activities.

Some methodological issues

One of the most important issues when preparing methodologies for new data collection activities is that of the use of established and acknowledged international classifications. This is especially the case when the new basic data are likely to be used for comparisons with other series of economic, industrial, foreign trade, manpower, etc. data, either for isolated analytical purposes or notably for building up new series of indicators.

One option is to build the new set of data entirely within an existing framework, for example the System of National Accounts (SNA). The choice for R&D statistics has been to develop outside SNA but to use United Nations classifications and concepts wherever these are appropriate.

Thus, the sectors of performance of R&D are broadly defined, in the "Frascati Manual", in terms of the SNA. Industrial R&D statistics are presented using the ISIC

(the International Standard Industrial Classification), and the broad fields of science classification used for the Higher Education and Private Non-Profit sectors follow ISCED (the International Standard Classification of Education). ISCED is also used to define categories of R&D personnel in terms of formal qualification in the Manual, whereas ISCO (the International Standard Classification of Occupations) serves for the parallel definition of R&D personnel by occupation/function.

The more sophisticated (i.e. the more disaggregated) the level at which comparisons between data bases are undertaken, the greater the problems. For reasonably accurate comparisons (between countries, different series of data, etc.) one may sometimes have to act at the level of the smallest common denominator which, in fact, may be at a rather aggregated level of detail. A good knowledge of the problems involved is therefore needed for any analytical studies based on such statistics and indicators.

However, the international classifications have not necessarily been developed for use in line with the purposes of the indicator development activities and some adaptations to specific needs may therefore be necessary, such as the definition of a separate Higher Education sector in the "Frascati Manual" (not separately available in SNA).

Special classification problems may arise where there are complex institutional differences between Member countries. For example, though it was agreed from the outset that the Higher Education sector should be separately defined in R&D surveys, variations in the way countries organise their academic and basic research have rendered it impossible to reach a clear definition of the institutional coverage of this sector. Problems of this kind may not be obvious when theoretical norms are first being set, but rather emerge over time as the data from international surveys are assessed and analysed.

Finally, for categories specific to the area concerned, it will be necessary to build entirely new classifications. For example, the "Frascati Manual" gives the definitive international definitions for separating R&D from other activities and for its breakdown between basic research, applied research and experimental development.

Data base work

STIID depends on Member governments to supply it with the data for inclusion in its data bases. The main method of obtaining such data is by means of questionnaires to Member countries which return data, using international norms, collected in domestic surveys. For the S&T indicators, the main sources are the biennial ISY (International Statistical Year) "benchmark" surveys, completed twice a year with questionnaires for the special data base segment of up-to-date indicators "MSTI" (Main Scientific and Technological Indicators) which in 1990 replaced the former data base segment "Rapid". Data are sometimes also drawn from national publications or from specific requests addressed to Member countries.

Once received and checked, the data are entered in the relevant "first-level" STIID data-base segments (usually available for "free distribution" according to the OECD rules for dissemination of data) and may then, for internal analytical purposes, be further processed and stored in the "second-level" part of the data-base ("restricted" or "private").

The larger data bases are stocked on the OECD Unisys A16 mainframe computer using two software systems developed for the Organisation: MEI (Main Economic Indicators), a time-series based data base system, and OSIRIS (OECD Statistical Infor-

mation Research and Inquiry System), a flexible multidimensional matrix-based system for the computerised storage and manipulation of large amounts of macro-economic statistical information.

In the case of R&D statistics/S&T Indicators, OSIRIS has proved to be a powerful and extremely user-friendly system, possibly thanks to the fact that the data series were from the outset well structured and, therefore, easy to handle also for non-experts, with a minimum of instruction and training. The system allows for data to be stored in multidimensional segments. (See Annex 2 for more information.)

Data base work is backed up by a suite of programmes for the entry and verification of data, extraction of tables, transformation, and preparation of graphics.

With a view to facilitating the access to the STIID data bases to outside users (principally non-STIID members of the OECD), an Inventory has been prepared to present the scope, the contents and the availability of data in the various "established" data segments.

Diffusion of data

General

Over the years the pattern of diffusion of S&T data has changed in line with consumer demand, new data- and word-processing capacities and the amount of Secretariat resources available for the task. Establishing a satisfactory policy and meeting the deadlines involved has not always been easy given that demand has usually exceeded supply.

In general, STIID diffusion policy is to provide the appropriate data as rapidly as possible for government policy-makers and those who work directly for them, and to supply information to a wider range of users sometimes less regularly and, increasingly, on payment. The market can be broadly divided into two categories: "main indicators" and "basic statistics".

Main science and technology indicators

It was clear from very early in our work that most potential users of S&T statistics were interested in a relatively small number of series and usually requested them already in the form of "indicators", of which the most popular is still R&D expenditure expressed as a percentage of GDP. The arrangements made to circulate a timely scoreboard of indicators have been gradually formalised and improved over time and the Division now maintains a special data base which, until 1989, contained 54 such indicators. From 1990 onwards, the number of indicators has been increased to 89.

The data base is updated twice a year and the results are published (together with summary methodological information) as "Main Science and Technology Indicators" for sale on paper and on diskette and immediately available to Member governments via the OECD's On-line Information Service (OLIS). (See Annex 3.)

The success of this data base and associated publications depends on rapid response from national experts to the OECD questionnaire. This has always been forthcoming as they are amongst data bases main users and have participated closely in its design.

The "Main Science and Technology Indicators" segment of the data base serves daily to meet a large number of internal and external requests for information, such as for OECD Science Policy reports, "parliamentary questions" from Member countries and demands from the press. Furthermore, in order to present the main results of its work to a wider audience than that of the "restricted users" mentioned above, the STIID publishes at irregular intervals a *Newsletter* containing most recent "highlights" and results of the ISY surveys or other activities.

Statistical information will also be found in the statistical sections of various STIID analytical publications such as the *OECD Science and Technology Indicators* report series (1984, 1986 and 1989).

Detailed basic statistics

The efficient diffusion of the detailed contents of the STIID data base of R&D statistics has proved to be more difficult for a number of reasons, notably: the questionnaires come in from Member countries spread over a long period; the mass of data is considerable now that long time-series are available and that quality control and annotation of all series to the level required for publication is very time-consuming.

As wider interest in S&T indicators revived, it became necessary to review the policy for diffusing detailed data. All the previous volumes have been discontinued. The whole of the first-level data base has been issued on magnetic tape for use by analysts who wish to make their own detailed studies. A biennial publication, to be entitled "S&T Basic Statistics", will be issued in 1991 and cover the most often requested basic statistics plus related methodological notes. These tables will also be issued in diskette form. Furthermore, in order to ensure that the new users (and potential purchasers) of OECD S&T statistics get appropriate advice as to their use and interpretation, a staff member has been nominated to deal with special requests for data. Still, there continues to be difficulty in meeting demand.

Analytical work

General

It is often argued that it is the job of a statistical service to collect, process and issue statistics but not to analyse them as the latter is the proper task of policy staff or of academics. This view has never been accepted in the case of OECD S&T indicators work. Certainly, policy staff and academics can and should undertake studies based on new indicators but these will usually be centred on answering one question or testing one hypothesis. There is also a need for more general international comparative studies based on the systematic analysis of data by those who have detailed knowledge of the problems of interpretation and comparability. The STIID has always undertaken and published such studies which are now issued regularly in the *OECD Science and Technology Indicators* reports.

These analytical studies have proved to be of extreme importance to test the utility and the reliability of the data already collected, and to discover items where the Frascati guidelines were inconsistent or otherwise unsatisfactory.

Analytical second-level data bases

The growing sophistication of computing resources has facilitated analysis but has also brought new challenges. The availability of user-friendly data-analysis and econometric packages means that it is now relatively easy to study longer and more complex sets of data. For example, it is now feasible to compare series of technological and industrial indicators simultaneously for 25 OECD countries, 40 industries over a period of 20 years, whereas in the past, cuts had to be made to obtain a "manageable" matrix.

The downside of this new capability is that all the data used must be comparable for all the variables, between countries and over time, and that the kind of *ad hoc* adjustments made for past analytical studies must be systematised and entered into new "second-level" data bases. Despite the use of sophisticated techniques to match series and fill gaps (in view of constructing all-OECD, all-EEC totals, etc.), the establishment of such second-level data bases involves a considerable investment in quality control, both at the OECD and by Member countries.

The principal analytical data base development currently under way in the STIID is that of "STAN", the STructural ANalysis data base, in which data from a series of OECD and other international industrial data bases (R&D, industrial production and manufacturing trade, etc.) are merged at a rather disaggregated level. STAN has been designed specifically to underpin analysis of the connection between technology, structural adjustment and economic performance in OECD countries and is intended to cover, in the somewhat longer term, the full data spectrum from basic science indicators through to trade and economic performance indicators. Powerful computer procedures are used to fill the gaps in "patching" data obtained from different sources.

The development of "STAN" is a complex and resource-consuming exercise, but one of great interest and potential both to the Secretariat and several of its Committees and to the Member countries. In the somewhat longer term, "STAN" may perhaps serve as a model for a similar approach with regard to the "academic" sector, incorporating data for inputs, throughputs and outputs (students, diplomas, employment, etc.) of national higher education systems, data for university research activities, and, possibly, various kinds of "output" indicators (bibliometrics data, for instance) of the same R&D activities.

VII. CONCLUSION

The development of new indicators takes a considerable amount of time and resources. As we have seen, different types of resources are required at different stages. Despite the invaluable facilities and opportunities offered by modern computers, there is no substitute for quality control by staff with a deep knowledge of the field concerned. This knowledge takes time to acquire and, once the investment in building a team of specialists has been made in national capitals and at the OECD, it is essential that it should be used productively. Hence, there is a need to undertake at regular intervals a review of any statistical programme to see what changes need to be made to current projects and to decide priorities for the future.

The OECD has had a long and detailed association with the development of indicators from a statistical base. The international S&T indicator system which already exists and its development can provide valuable lessons for those involved in the development of internationally comparable indicators and for educational policy-makers.

Annex 1

THE HISTORY OF THE "FRASCATI MANUAL"

Encouraged by rapid growth in financial and human resources devoted to scientific and technological research, most Member countries started the collection of some kind of statistical data in this field around 1960. Differences in scope/coverage, methods and concepts, however, made international comparisons difficult and countries encountered theoretical difficulties when starting their national R&D surveys. An increasing need was thus felt for some attempt at standardization of the kind already undertaken for economic statistics.

Already in 1957, the OEEC (OECD's predecessor) had begun to convene meetings of experts to discuss methodological problems in the field.

When the Directorate for Scientific Affairs – the predecessor of the present Directorate for Science, Technology and Industry (DSTI) – took over the OEEC work in 1961, the time was ripe for specific proposals for standardization. In 1962, it was decided to convene a workshop on the technical problems of measuring R&D.

The first "Frascati Manual"

A consultant, Mr. Christopher Freeman, was appointed by the Directorate to prepare a draft document which was circulated to Member countries and revised in the light of their comments. The proposed guidelines for the measurement of R&D were discussed, revised and accepted by national experts at a first conference held at Frascati (Italy) in 1963.

Later in 1963 the Directorate invited the UK National Institute for Economic and Social Research to undertake an experimental comparison of R&D efforts in five West European countries, the United States and the Soviet Union. The institute study, though based on statistics from surveys undertaken before the above-mentioned standards had been decided upon, also tested the first draft definitions. The report concluded that the available statistical information still left a great deal to be desired and suggested a series of improvements.

The first international R&D survey (ISY)

In 1964, following the acceptance of the "Frascati Manual", OECD launched its first "International Statistical Year (ISY)" survey of resources devoted to R&D. Seventeen countries took part in this survey and returned data for the year 1963 or 1964, many of them conducting national surveys and special enquiries for the first time.

Revisions of the Manual – the role of evaluation *ad hoc* groups

Following the publication of the first ISY results, the Committee for Science Policy requested the Secretariat to prepare a revision of the "Frascati Manual" in the light of the experience gained. A draft revision was examined at a second meeting of experts at Frascati in 1968.

During this revision, particular attention was paid to making the Manual as comparable as possible with existing United Nations' international standards such as the System of National Accounts (SNA) and the International Standard Industrial Classification (ISIC). The revised version of the Manual was published in 1970.

When work started on the third revision, Member countries had by 1973 participated in four ISY surveys and the accuracy and comparability of the data had much benefitted from this continued experience. National survey techniques had greatly improved.

In a situation with increasing budgetary constraints where Member countries' opinions as to the resources to be devoted to the R&D statistics varied widely, the Committee for Scientific and Technological Policy (CSTP) in 1972 decided to set up a first *ad hoc* R&D statistics review group to advise the CSTP and the Secretariat on how to make optimal use in the short term of the restricted resources available at the OECD while taking account of Member countries' priorities. Member countries were asked to dress up inventories of their needs and nearly all responded. In addition to giving absolute priority to a continuation of the ISY surveys, they made a number of recommendations touching on methodology, notably concerning the need for closer contacts between the OECD and other international organisations (such as UNESCO).

As a result, the third revision of the "Frascati Manual" went more deeply into those subjects already treated and examined certain new subjects. In response to a growing demand from science policy-makers and from members of the Group of National Experts on S&T Indicators (NESTI) the scope of the Manual was expanded to cover R&D also in the social sciences and humanities (SSH). Until then, the OECD data collection had been limited to the natural sciences and engineering only (NSE) whereas UNESCO already collected SSH data. Greater stress was given to "functional" classifications (such as the collection of R&D data by "socio-economic objectives"). The third edition of the Manual was adopted in 1974.

When work began on the fourth – and current – revision of the 1981 Manual, NESTI recommended that only an intermediate exercise should be undertaken, with no significant changes in key concepts and classifications. The main stress was on improving drafting and layout. (The earlier versions of the Manual had essentially been written for the attention of experts in Member countries who conducted the R&D surveys and collected and issued the national R&D series.) In order to facilitate the understanding of the data and their analysis to a wider audience, a new Chapter I – addressed principally to non-experts – was drafted.

However, a number of other revisions were made reflecting the considerations of a second CSTP review group (set up in 1976), of the Secretariat and of NESTI. The task of the new group was to assist the Secretariat with further medium- and long-term guidance on priorities, taking into account once again the needs of actual and potential users and the resources available both at the OECD and in Member countries. The general conclusions of this second group, which represented the users rather than the producers of R&D statistics, were very similar to those of its predecessor, giving highest priority to the ISY (input) surveys of financial and human resources devoted to R&D by Member countries.

On the methodological and analytical front, the Group recommended that the Secretariat should continue to work in the long term on the development of direct or proxy indicators of the results (output) of the R&D inputs.

Another interested party to this revision was the OECD Secretariat itself. Over the years, the Science and Technology Indicators Unit responsible for the R&D data collection activities had become a primary "consumer" of its own statistical series and other related data, by undertaking analytical studies either independently or within the framework of other OECD projects. This work

had shed additional light on problems of international comparability which obviously justified further clarifications in the Manual. In addition, efforts were made in the revised Manual to ensure that its theoretical guidelines should tie in as closely as possible with the practical recommendations in the ISY questionnaires.

The "Frascati Manual" supplement dealing with higher education R&D

Even if the Manual was thought to have reached general maturity with its fourth edition (at least as far as the input guidelines were concerned), one major problem still persisted, that of measuring the R&D resources in the Higher Education (HE) sector. In fact, this institutional entity is not a separate sector in the System of National Accounts (SNA) but had been separately identified in the early days of the "Frascati Manual" by the OECD (and also by UNESCO) because of the important role played by universities and similar institutions in performing the national R&D efforts and training scientific personnel.

These problems were due both to the institutional and to the functional aspects of university research. For various (administrative, historical, financial and other) reasons, organisational set-ups differ widely between countries. Given the close links between teaching, research and administrative activities (or, for instance, between specialised medical treatment and R&D in university hospitals) of faculty staff and of post-graduate students who are frequently involved in both research and teaching (at lower levels of the university system), international comparability of R&D statistics was found to be much lower in this sector than for most other series of R&D data.

Several NESTI sessions had been devoted to this problem over the years and in 1985 a seminar took place at the OECD to propose practical solutions, based on national experience, to improve the situation. Amongst other things, special attention was given to presentations of various kinds of "time-budget" surveys conducted in Member countries to estimate the R&D shares of total working time for university staff and methods to derive the corresponding R&D expenditures using the "R&D coefficients" drawn from the same surveys. On the same occasion, experts also had the opportunity to discuss quantitative aspects of the measurement of outputs of university (and/or basic) research, such as various bibliometric methods.

As a follow-up to this seminar, and after discussions at the regular NESTI meetings and by a small group of national experts, a Supplement to the "Frascati Manual" dealing with *R&D Statistics and Output Measurement in the Higher Education Sector* was published by the OECD in March 1989.

Drawing on the conclusions of the Technology-Economy Programme (TEP) during which needs for new and/or improved R&D/S&T indicators were identified, the next revision of the "Frascati Manual" is underway. A meeting was scheduled to take place in Italy in the Autumn of 1991. This fifth revision will also have to take into account the consequences of recent or ongoing revisions of some of the principal world level classifications (SNA, ISIC, ISCO, etc.).

Annex 2

TYPES OF R&D DATA AVAILABLE IN THE STIID "FIRST-LEVEL" DATA BASE

By sectors of performance

R&D expenditures and personnel data are collected for the following four OECD sectors of R&D performance: the Business Enterprise sector, the Government sector, the Higher Education sector and the Private Non-profit sector, the resources of which, taken together, make up the national totals (with the exception of the Higher Education sector, this sectoring differs from that used by UNESCO in their R&D surveys).

Expenditures

Expenditures data are collected by "type of cost" and by "source of finance" (of the R&D performed). Types of cost data are broken down into "current expenditures" (salaries and other personnel costs on the one hand, and other current expenditures on the other) and "capital expenditures" (for land and buildings, and instruments and equipment, respectively). For the sources of finance, the same four sectors as those described above are used, completed, however, with a fifth sector ("from abroad"). Depending on the sector, there are also a few additional breakdowns such as (performers') own funds and, – in the case of Higher Education sector R&D – a supplementary breakdown of "public finance" into direct Government funds and General University Funds (GUF), separating the ear-marked public R&D funds from the block-grant type of public finance.

Personnel

The R&D personnel series are collected in full-time equivalence (FTE) rather than on a head-count basis. Two separate breakdowns are requested in the questionnaires, one by "occupation" (defined in terms of ISCO), the other by "formal qualification" (defined in terms of ISCED). The former asks for FTE data for research scientists and engineers (RSE), technicians and other categories of R&D personnel; the second for data broken down by holders of university-level degrees, of other post-secondary diplomas, of diplomas of secondary education, and "other" qualifications. Not all countries report data for both personnel classifications but usually for only one of the two; for the purposes of estimating various totals (OECD, EEC, etc.), the categories of RSE and university diplomas are usually considered as proxies for one another.

Institutional sector breakdowns

Both the expenditure and R&D personnel series are classified according to institutional breakdowns of the sectors of performance concerned. The Business Enterprise sector series are broken down according to an ISIC-related industrial classification; data for the Higher Education and the Private Non-profit sectors according to a breakdown by broad (ISCED) fields of science. The former institutional breakdown of the Government sector was abolished by the middle of the 1970s following *ad hoc* review group recommendations considering that the international comparability of the data reported was unsatisfactory; NESTI experts have not as yet been able to agree on another acceptable compromise classification.

Annex 3

OECD/STIID MAIN SCIENCE AND TECHNOLOGY INDICATORS

1. Gross Domestic Expenditure on R&D – GERD (million national currency)
2. GERD (million current ppp $)
3. GERD – Compound annual growth (fixed prices)
4. GERD per capita population (curent ppp $)
5. GERD as a percentage of GDP
6. Estimated Civil GERD as a percentage of GDP

7. Total R&D personnel (FTE)
8. Total R&D personnel – Compound annual growth
9. Total R&D personnel per thousand labour force

10. Total Research Scientists and Engineers – RSE (or university graduates) (FTE)
11. Total RSE (or university graduates) – Compound annual growth
12. Total RSE (or university graduates) per thousand labour force

13. Percentage of GERD financed by industry
14. Percentage of GERD financed by government
15. Percentage of GERD financed by other national sources
16. Percentage of GERD financed by abroad

17. Percentage of GERD performed by the Business Enterprise sector
18. Percentage of GERD performed by the Higher Education sector
19. Percentage of GERD performed by the Government sector
20. Percentage of GERD performed by the Private Non-Profit sector

21. Business Enterprise Expenditure on R&D – BERD (million national currency)
22. BERD (million current ppp $)
23. BERD – Compound annual growth (fixed prices)
24. BERD as a percentage of DPI
25. BERD as a percentage of GDP

26. Total Business Enterprise R&D personnel (FTE)
27. Total Business Enterprise R&D personnel – Compound annual growth
28. Total Business Enterprise R&D personnel as a percentage of national total

29. Business Enterprise RSE (or university graduates) (FTE)
30. Business Enterprise RSE (or university graduates) – Compound annual growth
31. Business Enterprise RSE (or university graduates) as a percentage of national total

32. Percentage of BERD financed by industry
33. Industry-financed BERD – Compound annual growth (fixed prices)
34. Industry-financed BERD as a percentage of DPI
35. Percentage of BERD financed by government

36. Percentage of BERD financed by other national sources
37. Percentage of BERD financed by abroad

38. Percentage of BERD performed in the aerospace industry
39. Percentage of BERD performed by the electrical/electronic industry
40. Percentage of BERD performed in the office machinery and computer industry
41. Percentage of BERD performed in the drug industry
42. Percentage of BERD performed in all other manufacturing industries
43. Percentage of BERD performed in non-manufacturing industry

44. Higher Education Expenditure on R&D – HERD (million national currency)
45. HERD (million current ppp $)
46. HERD – Compound annual growth (fixed prices)
47. HERD as a percentage of GDP
48. Higher Education total R&D personnel (FTE)
49. Higher Education total R&D personnel – Compound annual growth

50. Higher Education RSE (or university graduates) (FTE)
51. Higher Education RSE (or university graduates) – Compound annual growth
52. Higher Education RSE (or university graduates) as a percentage of national total

53. Government Intramural Expenditure on R&D – GOVERD (million national currency)
54. GOVERD (million current ppp $)
55. GOVERD – Compound annual growth (fixed prices)
56. GOVERD as a percentage of GDP

57. Government Total R&D personnel (FTE)
58. Government Total R&D personnel – Compound annual growth

59. Government RSE (or university graduates) (FTE)
60. Government RSE (or university graduates) – Compound annual growth
61. Government RSE (or university graduates) as a percentage of national total

62. Total Government Budget R&D Appropriations – GBAORD (million national currency)
63. Total GBAORD (million current ppp $)

64. Defence budget R&D as a percentage of Total GBAORD
65. Civil budget R&D as a percentage of Total BGAORD
66. Economic Development programmes as a percentage of Civil GBAORD
67. Health and Environment programmes as a percentage of Civil GBAORD
68. Space programmes as a percentage of Civil GBAORD
69. Non-Oriented Research programmes as a percentage of Civil GBAORD
70. General University Funds (GUF) as a percentage of Civil GBAORD

71. National patent applications
72. Domestic patent applications
73. Foreign patent applications
74. External patent applications

75. Foreign patent applications/Domestic patent applications ratio
76. Domestic patent applications/National patent applications ratio
77. National patent applications per 10 000 population
78. Rate of diffusion of patents

79. Technological balance of payments: receipts (million national currency)
80. Technological balance of payments: payments (million national currency)
81. Technological balance of payments: balance (million national currency)
82. Technological balance of payments: coverage ratio
83. Technological balance of payments: total transactions (million national currency)

84. Export/Import ratio: Aerospace industry
85. Export/Import ratio: Electrical/Electronic industry

86. Export/Import ratio: Office machinery and computer industry
87. Export/Import ratio: Drug industry
88. Export/Import ratio: Other manufacturing industries
89. Export/Import ratio: Total manufacturing

Abbreviations

R&D Terminology:

BE	Business Enterprise (Industry)
BERD	Expenditure on R&D in the Business Enterprise Sector
FTE	Full-time Equivalent (on R&D)
GBAORD	Government Budget Appropriations or Outlays for R&D
GERD	Gross Domestic Expenditure on R&D
GOVERD	Government Intramural Expenditure on R&D
GUF	General University Funds
HE	Higher Education
HERD	Expenditure on R&D in the Higher Education Sector
PNP	Private Non-Profit Sector
R&D	Research and Experimental Development
RSE	R&D Research Scientists and Engineers

Other:

DPI	Domestic Product of Industry
GDP	Gross Domestic Product
PPP	Purchasing Power Parities
TBP	Technology Balance of Payments

Chapter 6

STANDARDIZING DATA IN DECENTRALISED EDUCATIONAL DATA SYSTEMS

by

Ramsay W. Selden
Council of Chief State School Officers
United States

I. INTRODUCTION

To be useful, indicators of educational quality must be valid, accurate and comparable. Comparability is difficult to achieve in statistics gathered through decentralised statistical systems. Decentralised systems are those such as international or federal systems, where individual nations, states, or other entities have autonomy and responsibility for developing statistical systems. In decentralised systems, statistical programmes and procedures used by each reporting unit to collect statistics usually have unique characteristics. The system used by each reporting unit, such as a school, state, or nation, will have developed historically with distinct features to meet the specific needs of the country or jurisdiction. Survey forms, the timing and methods of data collection will vary among reporting units in the system.

The result of this variation is that statistics pertaining to "schools", to "enrolment", to "secondary completion", or other data elements, while seeming to involve similar terms and procedures, may be very different.

This chapter describes a systematic process by which the comparability of statistics collected in decentralised systems can be studied and improved. The experience reported here, while pertaining to efforts to improve comparability of statistics among states in the United States, can be extended to other situations as well, including international educational statistics.

II. THE PROJECT

This approach to improving data comparability was developed through the Education Data Improvement Project conducted by the Council of Chief State School Officers for the US National Center for Education Statistics (NCES) [1]. The project concentrated its efforts on NCES's Common Core of Data (CCD). These are data on public schools, including identification number, address, type and enrolment; secondary school gradu-

ates; school staff; and revenues and expenditures. Data pertaining to schools are reported on a universal basis for all public schools in the nation. Other data such as staffing, graduates and fiscal variables are reported as state aggregates. The project had three goals: to review the comprehensiveness and coverage of data in the Common Core; to study and improve the comparability of data reported among states; and to recommend changes to improve such features as the comparability, accuracy and timeliness of the data. Here, we describe only work on comparability, but various reports from the project address the other goals.

Any approach to improving data comparability in a decentralised system should involve the following elements or steps:

- empirical analysis of comparability;
- direct confrontation of comparability problems;
- recommendation of standard definitions to address problems and ensure comparability;
- establishment of standards for data collection;
- development of guidelines for reporting;
- implementation of definitions and procedures;
- monitoring of reporting against standards; and,
- enforcement and provision of incentives for compliance.

Together, these elements constitute an approach or paradigm for improving data comparability in a decentralised statistical system.

Empirical analysis of comparability problems

The first step was detailed analysis and documentation of comparability problems.

In the United States, it was widely presumed that definitions and procedures used by states in reporting the Common Core of Data differed and resulted in non-comparable data, but this had never been determined systematically. Establishing that there are problems in comparability and describing those problems specifically are essential, otherwise the problems cannot be confronted effectively.

To analyse the comparability of state data reported in the Common Core of Data (CCD), we conducted an exhaustive semantic content analysis of the forms and procedures used. We conducted what we called a "shuttle" survey, starting by assembling each state's data-collection forms and the guidebooks used to collect CCD data elements. We then reported to states what we understood their definitions and procedures to be. We asked state staff to confirm these understandings and we went back and forth with staff in the states to clarify and specify in exhaustive detail how their data elements might differ from, or be similar to, other states.

The method for such an analysis is to identify a comprehensive set of the semantic components that define each data element, to array these definitional and procedural components for each data element, and then to code the presence or absence of each component of the definition for each state (or other reporting unit). Does the state incorporate that component or not in its definition or procedures?

Two examples of the results of this process are shown in the following pages. The first is each state's standing on an exhaustive set of elements constituting a definition of "enrolment" in the Common Core of Data. Across the top are components of the

concept of enrolment, encompassing who is counted, when counts are taken, and whether counts are cumulative or residual. Considerable variation can be seen among the states in how these various components of a definition of enrolment are used and in how data are defined and collected. Some of these definitional differences resulted in numerical differences of some magnitude: one state included pre-school enrolments which another state did not count, resulting in a shift of several hundred thousand children.

The second example shows elements present or not in states' definitions of "public schools". Here, one of the inconsistencies had a quantitative effect of about 10 per cent: 10 per cent of the entities counted as schools in one state would not have been counted in another state.

In any decentralised statistical system, the comparability of definitions and other procedures affecting the substance of a data element must be studied systematically, to determine whether data reported by reporting units are comparable and to determine where, specifically, they are not comparable. As will be explained below, this is an essential first step in addressing areas of consistency so that comparable data can be obtained.

Confrontation of comparability problems

The second step in achieving comparability is confronting the problems that are discovered. Doing this is much easier if the problems have been laid out concretely and inescapably.

After analysing and describing the comparability problems, we convened meetings of representatives of the states' educational data programmes to review the results. These are the professional staff who design and maintain the states' data collection and computer information systems in education. At these meetings, we reviewed the results of the comparability analysis, but we made sure that there were three types of people present: "old guard", conservative representatives of state data systems; "progressive", representatives; and experts in the type of statistics we would be discussing: enrolment data, drop-out statistics, fiscal data, and so on.

The groups were composed in this way to enable three things to happen. The old guard would have an opportunity to explain why they had been collecting statistics the way they had for so long according to their own definitions and procedures, and they could express how difficult it would be to change to new, standard definitions and procedures. The progressives could explain how important it would be to change over to comparable definitions, and they could volunteer on behalf of their states to change over in the interest of obtaining high-quality data to allow sound and meaningful comparisons among states. The technical experts were on hand to explain how it might be technically proper and practically feasible to change over to valid, consistent definitions. Thus, we were able to confront the problems in a non-threatening way and to give states an opportunity to alleviate these problems as a result of their own professional initiative.

Recommending standard definitions and procedures

Once problems were confronted, we could proceed to come up with definitions and data collection procedures that would assure standardized, comparable data.

Definitions and procedures were developed by technical task forces consisting of state and local data collection specialists and technical experts in each particular area of data collection. They built upon the findings of the comparability studies and attempted to specify components of the definitions in order to resolve variations or ambiguities in the definitions among the states. The goal was a valid, complete, and unambiguous definition and set of collection procedures for each data element, such as enrolment, school, graduate, expenditure, or teacher that would ensure acceptably comparable data in the future. National experts were convened to make sure that the definitions would not just represent a compromise or majority position among the practices of the states, but would represent the thinking about that concept and how it should be defined validly and collected accurately.

For each data element, draft definitions and proposed data collection procedures were circulated among states and technical committees for comment. Comments were considered and appropriate revisions were made accordingly. States were very carefully given opportunities to review and comment on the definitions and procedures before they were adopted and, ultimately, the heads of the education systems in each state were given an opportunity to consider and approve the definitions. This was important, because in some cases the state commissioners and superintendents cared specifically about data collection requirements or changes that might be imposed on their states, and later we knew it would be important to be able to call on support from the highest levels to reinforce the implementation of the new statistics and definitions.

III. ESTABLISHMENT OF STANDARDS FOR REPORTING

Once definitions and procedures are recommended and accepted by the states, they can be translated into standards for data providers, by which submissions of data can be evaluated for comparability.

The definitions and procedures recommended in this project were submitted to the National Center for Education Statistics (NCES) for incorporation into the procedures and reporting guidelines for the Common Core of Data. At that time, NCES was establishing general standards for statistical programmes, encompassing response rates for surveys, precision of measurement instruments, and other matters. Among the general standards developed by NCES are some provisions for data element definitions and instructions for completing surveys. Combined with the general standards, the specific, concrete definitions established by this project for the CCD constituted standards by which submissions by the states could be evaluated for comparability and adherence to uniform definitions. A concrete reference point was available by which submissions from states could be reviewed for their consistency.

Since the project, our Council has begun reporting comparative statistics on education among the states[2]. Before we were willing to report enrolment figures from the states, we asked for assurance from NCES programme staff that the figures were comparable and consistent with the definitions our project had established.

The process of applying definitions and procedures to develop standards can be done more deliberately and explicitly. In most decentralised data collection systems, such as international data collection programmes, this would be desirable. Once definitions and procedures have been agreed on, they would be transformed into specifications, targets,

and ranges within which allowable data can fall, and these would be communicated as part of the data collection activity to those providing the data. Examples of allowable and non-allowable departures from standard definitions could also be given.

In fact, in our project the definitions are being used implicitly as standards by the programme office within NCES which administers the Common Core of Data. They are communicated to data providers as standards by which submissions will be reviewed through their incorporation into guidelines for reporting, and NCES awards professional recognition each year to states adhering to the new definitions and procedures.

Development of guidelines for reporting

For uniform, unambiguous definitions to be adopted and for standards to be enforceable, guidelines are needed to specify to data providers how to report.

These guidelines must be complete: they must cover all aspects of the data element definition or data collection procedures into which ambiguity or lack of clarity may enter. They must be effective, defining the data element and resolving potential ambiguities. They must provide realistic directions and solutions. They must be clear and concrete, explaining to data providers what is meant and intended. They should also be simple and direct without compromising their effectiveness in specifying the data element. And they should be usable in a format conducive to use by data providers, and being readily available, credible, and authoritative.

Implementation of new definitions and procedures

After new definitions and procedures for data collection are developed, they must be implemented by the reporting units in the data collection programme – states, countries, or whatever.

For some reporting units, the newly standardized definitions and procedures will present no real problems: the new standards will be consistent with definitions and procedures already being used. For others, changes will be necessary if they are to provide data consistent with the new definitions. The responses of reporting units (states, countries) might take on one of the following forms, though no one system will be readily suitable to every reporting unit:

- Change their data systems to the new definitions and procedures;
- Maintain their old systems while initiating a new, parallel system to supply data according to the new definitions and procedures;
- Convert or cross-walk their data to new definitions and formats, if possible;
- Supply marginally comparable data where they are available; or
- Decline to submit data.

Because of the difficulty of states or countries converting to new standards and definitions, time, patience, support, and technical assistance should be provided to facilitate the process. In order to speed the process, attention from the highest levels of political management in the state or country should be obtained. Often, data system managers struggle in obscurity, trying slowly to change their systems without the attention or support of their leaders to speed the process along and provide resources within their organisations to make the changes. If data collection programmes are high in public

and political interest, and inter-state and international comparisons often are, this interest can be used to direct organisational support towards revising the data system.

Monitoring adherence to standard definitions and procedures

Once definitions, standards, and guidelines are available, submissions by data providers can be reviewed regularly in an ongoing way to ensure comparability. Depending on the resources available to the data collection organisation, this can be done by:

- Interviewing staff of the submitting organisations;
- Conducting periodic comparability checks, perhaps on a spot-check basis, replicating in mini-format the comparability analysis described in Section II as our first step;
- Securing an occasional third-party study to check the comparability of the data submitted.

Some sort of monitoring or verification of adherence to standard definitions and guidelines is necessary to assure the programme of high-quality data that can be compared legitimately among reporting units and to maintain pressure on reporting units to comply with definitions and standards. Monitoring can also maintain pressure on a reporting unit to come into compliance if they have not done so.

Ultimately, in order to ensure comparability of data, an educational data system must have some principle by which uniform definitions and standards are enforced. This is especially true for a decentralised system, where the system is dependent on data providers to comply with standards. In centralised or integrated systems, more direct control over data collection processes is possible.

Some of the incentives or mechanisms for obtaining compliance with definitions and standard procedures are:

- non-reporting of non-complying data;
- reporting all data, but distinguishing or noting non-complying data;
- linking compliance to other organisational benefits; or
- directing resources to reporting units (states, countries) experiencing difficulty in reporting.

Non-reporting of non-complying data can be a powerful incentive when the data generate a great deal of interest by the public or policy-makers, such as inter-state or international comparisons. Noting non-complying data is weaker, but still exerts pressure by indicating whose data may be of questionable comparability. In some systems, organisational benefits, such as fiscal aid, are tied to participation in data systems. This can be considered when the organisation operating the data collection programme provides significant support to its constituents and where there is the political will to emphasize data collection in this way. Directing resources to those experiencing problems may be the most constructive professional response, addressing the technical and programmatic reasons for not reporting.

IV. CONCLUSION

The development of indicators that can be compared across systems and countries is bedevilled by the need to retain accuracy and meaning.

In this chapter, a process for studying and improving data comparability in decentralised data systems was reported. This instance was the state-federal data system in the United States and several specific steps or elements in the approach were outlined, with suggestions or descriptions of how each of them can be conducted.

This approach could be applicable to federal data systems in other countries and stands as an interesting case study in the development of international education indicators.

NOTES

1. NCES Contract No. 300-85-0146, United States Department of Education.
2. Council of Chief State School Officers, *State Education Indicators*, Washington, D.C. (CCSSO) 1987, 1988.

Chapter 7

WHAT CAN BE LEARNED FROM ALL THIS?

by

Isabelle Delfau
OECD/CERI Secretariat

and

Ramsay W. Selden
Council of Chief State School Officers
United States

Looking back at the first eighteen months of active and sometimes hurried work on the International Education Indicators Project, there are a few lessons and experiences in international planning and work which deserve consideration.

First is the importance of consensus and decentralisation as key work processes. Twenty-four Member countries participated actively in the Indicators Project, and a simple option would have been to have made decisions on a majority basis. Indeed, majority rule is often taken for granted in international work. Accepting it ensures that a decision is always reached no matter how difficult the preceding debate. But while this undoubtedly saves time in the short run, it may be to the detriment of the long-term interests of a programme. The OECD Indicators Project was designed as a long-term effort to construct an international information system on education. For such an endeavour to succeed, participants must feel totally comfortable with what is being developed and the elements that go into it. There is otherwise a risk that support may not be forthcoming, if the product is not deemed useful in terms of national needs and priorities. The Project departed from traditional approaches by working from the bottom up to ensure that the indicators developed matched education authorities' expectations and it was insisted that consensus must be reached on each element considered for reporting. This proved difficult, exciting and worthwhile. The approach combined a commitment to decentralisation, devolution and involvement in administration with a co-operative division of work to maximise the effectiveness of available resources and expertise. Consensus was never mistaken for the right to veto, but was rather understood as the need to work jointly towards constructive compromises acceptable to all. This was a highly positive experience and ensured that the work coming from the groups participating in the Project, while being internationally meaningful, was deeply rooted in the various national realities and contexts.

The second lesson to be learned from the Project is the value of information sharing. For many of the participants, involvement in the Project meant first and foremost getting access to information on other systems, to international expertise and to different perspectives on familiar problems. Education systems are anything but accessible; information is there for those who really want it, but getting to it usually requires work and serendipity. The construction of Networks such as those seen in this Project and the concomitant development of professional relations can, at least in part, reduce the element of "luck". Getting the necessary information will always require work, but creating platforms and channels designed to ease and accelerate its flow ensures that the same amount of effort can be more efficiently targeted and yield higher benefits. For this fully to happen, the active co-operation of the OECD Secretariat and of Member countries is needed: the latter to provide the information, the former to facilitate access to it.

Third, there is the issue of coherence. From this perspective, a devolved structure is a mixed blessing. Its advantages cannot be contested: ensuring that the work is driven by those who will be its ultimate users and therefore know what is required; focusing the task to be discharged through medium-sized groups able to explore their domain freely; organising volunteer participation according to priorities and expertise. The costs cannot be minimalised: the price can be high in terms of needing to impose or overlay co-ordination and coherence. Coherence is not highly compatible with devolution, and it cannot be arrived at simply. It can only be the product of conscious and co-ordinated efforts: in offering a conceptual frame of reference that all groups can adhere to; in designing common instruments to report on the outcomes of the work; and in ensuring that work pursues shared aims. Thus, in the case of the Indicators Project, all Networks could relate to the model developed as an organising frame; a common "template" stressing the definition, rationale and data aspects was used to present the key indicators proposed for consideration; and a Scientific Advisory Group met regularly with the OECD Secretariat to overview work progress in light of the foreseen outcomes and to report to the Networks. The Advisory Group brought recommendations which had gone astray back to the organising scheme. Without due care to such matters, there is great risk in devolved work that there will be an impressive and intellectually fascinating firework of ideas, but one that proves ultimately difficult to form into anything that is manageable or coherent.

Fourth, the Project highlighted the necessity for interaction between Networks and sound linkages among the indicators developed. Indicators sometimes address relatively simple concepts which are directly measurable. More abstract concepts such as efficiency or productivity may require more complex indicators which have the advantage of compacting and articulating in one concept several aspects of the education system. Given the complexity and the diversity of education systems, it is obvious that individual and isolated indicators, and even compound indicators, can only convey limited information. It is therefore essential to build a coherent system of indicators that clearly and comprehensively reflects the state of an education system. Such a system cannot simply comprise a collection of fragmented statistical data, however easily accessible such data may be. Indicators that are devised without a clear overall concept are likely to produce a jumble of disconnected facts and figures. Building a coherent system or set of indicators is a way to compensate for the unidimensional limits of each indicator and to change an *ad hoc* collection of statistics into a valid representation of the condition of a system. To provide this picture, indicators should be selected so as to be logically linked, and these linkages should proceed from a model that describes how the education system works.

Having a reference model permits a broader assessment of an indicator's relevance; it is a crucial prerequisite to proper and valid interpretation and use of indicators. The key role of a conceptual frame is therefore to expand the interpretative value of the set of indicators to be developed. Ideally, such a frame will help observers understand how individual components of the set work together to achieve a given effect: labour market destination indicators and achievement indicators can, for example, help policy-makers put into perspective educational spending and cost indicators. The set can thus be used to assess current conditions, appreciate the likely effects of changes in policy and provide clues about what is wrong and what changes might lead to improvement.

In this Project, an interaction occurred between the individual indicators proposed by the Networks and the overall organising scheme. Over the course of the Project, the model caused indicators to be conceived to fill gaps, and the creativity and idiosyncratic nature of the Networks tested and expanded the model. Also, the ultimate use of the indicators was tested: the model implied a diagnostic interpretation of causes and effects, but given the associational approach which was adopted within the limits of some twenty countries, such causal interpretations could not be verified.

The fifth lesson of the Project is that the reference model needs strong and solid foundations. A very simple model of education environment, resource, process and output – was used throughout the exploratory phase of the Project and has been presented in this publication. It is obvious that such a model does not readily achieve the objectives identified above as desirable and must be refined to reach greater interpretative value. This implies accounting in the model for important elements currently missing, such as school level and attitudinal indicators, and exploring some of the theoretical or conceptual issues that were barely touched upon during the exploratory phase. It requires striving to develop a more dynamic model, one that is able to describe for instance how resources are allocated to education and to specific programmes, and how this allocation can be related to processes. On a different level, it also means that the greatest attention should be paid to the issues of presentation and interpretation of the information and the indicators ultimately reported: the importance of understanding the national contexts of the systems compared can hardly be over-stressed. Contexts include culture, history and values as much as structures and resources. Refining the model undoubtedly means facing difficulties of comprehensiveness and comparability, in addressing some of the many policy issues raised by the use of indicators as monitoring instruments: where does equity fit in? How can differences in school performance be related to differences in the school's environment or resources? How does the model relate to the debate on quality and/or relevance of education? Which variables can be seen to affect teaching quality, and how is it to be defined anyway? These questions are part of the changing background that shapes the reference frame used in educational research and thinking; as such, they cannot be left untouched.

Sixth, there is an unavoidable trade-off between data comparability and availability of the indicators. Information and data collected and used in the Project must be reliable and internationally comparable, but transforming basic national education statistics into a set of internationally comparable indicators is a complicated process. The Project and the Networks benefitted in this respect from the joint questionnaire that the OECD, UNESCO and the EC use for the collection of education statistics. Even this instrument is far from solving all the definitional problems involved in ensuring comparability of the data. The questionnaires are based on an internationally agreed-upon classification of education called ISCED – International Standard Classification of Education. ISCED was elabo-

rated over a twenty-year period, adopted in 1978 and used for the first time in 1983. While the existence of ISCED is a major step in international co-operation and harmonization of data collection in education, many problems still remain unresolved. Chief among these is the fact that ISCED is a classification of programmes of study grouped by level and type of education. As a consequence, it is difficult to derive from it information on teaching staff and expenditure and relate it to pupils and courses. It should also be noted that the twenty years it took to develop ISCED saw many major changes in educational policies and structures which are not reflected in the definitions finally agreed on and used. Even where ISCED exists and can be used, definitional problems remain and affect the comparability of data. Agreeing on the importance of an indicator such as pupil/teacher ratio, for instance, raises many questions as to what exactly is a teacher: part-time teachers have to be accounted for, teachers' assistants should either be considered as teachers or not, and this must be done consistently. Indicators which are not comparable or accurate are not credible. With the first serious challenge of a figure, a set of indicators can come tumbling down and potential users lose interest. There is however a point beyond which the search for exact comparability yields diminishing or even negative returns and can hinder constructing some indicators. Slight mismatches in definitions or data should be accepted, provided they are duly documented and do not radically alter the significance of the information reported. This should not preclude the indicator from being presented, with its explanatory footnotes and comments.

This publication attempted to bring out some of the most striking elements of the initial work undertaken by the Indicators Project. The exploratory phase was in many ways an enriching experience for most participants, and it is probably unfair to limit the experience to the important but far from all-encompassing dimensions discussed here. Together though, these dimensions seem worth setting down as interim landmarks. As they shape the ongoing work, it is important to record why and how some characteristics, some processes – a culture – came about. Possibly they will help others learn from the difficulties encountered by the Project, so that they can overcome their own all the more quickly.

MAIN SALES OUTLETS OF OECD PUBLICATIONS
PRINCIPAUX POINTS DE VENTE DES PUBLICATIONS DE L'OCDE

ARGENTINA – ARGENTINE
Carlos Hirsch S.R.L.
Galeria Güemes, Florida 165, 4° Piso
1333 Buenos Aires Tel. (1) 331.1787 y 331.2391
Telefax: (1) 331.1787

AUSTRALIA – AUSTRALIE
D.A. Book (Aust.) Pty. Ltd.
648 Whitehorse Road, P.O.B 163
Mitcham, Victoria 3132 Tel. (03) 873.4411
Telefax: (03) 873.5679

AUSTRIA – AUTRICHE
Gerold & Co.
Graben 31
Wien I Tel. (0222) 533.50.14

BELGIUM – BELGIQUE
Jean De Lannoy
Avenue du Roi 202
B-1060 Bruxelles Tel. (02) 538.51.69/538.08.41
Telefax: (02) 538.08.41

CANADA
Renouf Publishing Company Ltd.
1294 Algoma Road
Ottawa, ON K1B 3W8 Tel. (613) 741.4333
Telefax: (613) 741.5439

Stores:
61 Sparks Street
Ottawa, ON K1P 5R1 Tel. (613) 238.8985
211 Yonge Street
Toronto, ON M5B 1M4 Tel. (416) 363.3171

Les Éditions La Liberté Inc.
3020 Chemin Sainte-Foy
Sainte-Foy, PQ G1X 3V6 Tel. (418) 658.3763
Telefax: (418) 658.3763

Federal Publications
165 University Avenue
Toronto, ON M5H 3B8 Tel. (416) 581.1552
Telefax: (416) 581.1743

CHINA – CHINE
China National Publications Import
Export Corporation (CNPIEC)
P.O. Box 88
Beijing Tel. 403.5533
Telefax: 401.5664

DENMARK – DANEMARK
Munksgaard Export and Subscription Service
35, Nørre Søgade, P.O. Box 2148
DK-1016 København K Tel. (33) 12.85.70
Telefax: (33) 12.93.87

FINLAND – FINLANDE
Akateeminen Kirjakauppa
Keskuskatu 1, P.O. Box 128
00100 Helsinki Tel. (358 0) 12141
Telefax: (358 0) 121 4441

FRANCE
OECD/OCDE
Mail Orders/Commandes par correspondance:
2, rue André-Pascal
75775 Paris Cedex 16 Tel. (33-1) 45.24.82.00
Telefax: (33-1) 45.24.85.00 or (33-1) 45.24.81.76
Telex: 620 160 OCDE

OECD Bookshop/Librairie de l'OCDE :
33, rue Octave-Feuillet
75016 Paris Tel. (33-1) 45.24.81.67
(33-1) 45.24.81.81

Documentation Française
29, quai Voltaire
75007 Paris Tel. 40.15.70.00

Gibert Jeune (Droit-Économie)
6, place Saint-Michel
75006 Paris Tel. 43.25.91.19

Libraire du Commerce International
10, avenue d'Iéna
75016 Paris Tel. 40.73.34.60

Libraire Dunod
Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75016 Paris Tel. 47.27.18.56

Libraire Lavoisier
11, rue Lavoisier
75008 Paris Tel. 42.65.39.95

Libraire L.G.D.J. - Montchrestien
20, rue Soufflot
75005 Paris Tel. 46.33.89.85

Libraire des Sciences Politiques
30, rue Saint-Guillaume
75007 Paris Tel. 45.48.36.02

P.U.F.
49, boulevard Saint-Michel
75005 Paris Tel. 43.25.83.40

Libraire de l'Université
12a, rue Nazareth
13100 Aix-en-Provence Tel. (16) 42.26.18.08

Documentation Française
165, rue Garibaldi
69003 Lyon Tel. (16) 78.63.32.23

GERMANY – ALLEMAGNE
OECD Publications and Information Centre
Schedestrasse 7
D-W 5300 Bonn 1 Tel. (0228) 21.60.45
Telefax: (0228) 26.11.04

GREECE – GRÈCE
Libraire Kauffmann
Mavrokordatou 9
106 78 Athens Tel. 322.21.60
Telefax: 363.39.67

HONG-KONG
Swindon Book Co. Ltd.
13-15 Lock Road
Kowloon, Hong Kong Tel. 366.80.31
Telefax: 739.49.75

ICELAND – ISLANDE
Mál Mog Menning
Laugavegi 18, Pósthólf 392
121 Reykjavik Tel. 162.35.23

INDIA – INDE
Oxford Book and Stationery Co.
Scindia House
New Delhi 110001 Tel (11) 331.5896/5308
Telefax: (11) 332.5993

17 Park Street
Calcutta 700016 Tel. 240832

INDONESIA – INDONÉSIE
Pdii-Lipi
P.O. Box 4298
Jakarta 12042 Tel. 583467
Telex: 62 875

IRELAND – IRLANDE
TDC Publishers – Library Suppliers
12 North Frederick Street
Dublin 1 Tel. 74.48.35/74.96.77
Telefax: 74.84.16

ISRAEL
Electronic Publications only
Publications électroniques seulement
Sophist Systems Ltd.
71 Allenby Street
Tel-Aviv 65134 Tel. 3-29.00.21
Telefax: 3-29.92.39

ITALY – ITALIE
Libreria Commissionaria Sansoni
Via Duca di Calabria 1/1
50125 Firenze Tel. (055) 64.54.15
Telefax: (055) 64.12.57

Via Bartolini 29
20155 Milano Tel. (02) 36.50.83

Editrice e Libreria Herder
Piazza Montecitorio 120
00186 Roma Tel. 679.46.28
Telefax: 678.47.51

Libreria Hoepli
Via Hoepli 5
20121 Milano Tel. (02) 86.54.46
Telefax: (02) 805.28.86

Libreria Scientifica
Dott. Lucio de Biasio 'Aeiou'
Via Coronelli, 6
20146 Milano Tel. (02) 48.95.45.52
Telefax: (02) 48.95.45.48

JAPAN – JAPON
OECD Publications and Information Centre
Landic Akasaka Building
2-3-4 Akasaka, Minato-ku
Tokyo 107 Tel. (81 3) 3586.2016
Telefax: (81 3) 3584.7929

KOREA – CORÉE
Kyobo Book Centre Co. Ltd.
P.O. Box 1658, Kwang Hwa Moon
Seoul Tel. 730.78.91
Telefax: 735.00.30

MALAYSIA – MALAISIE
Co-operative Bookshop Ltd.
University of Malaya
P.O. Box 1127, Jalan Pantai Baru
59700 Kuala Lumpur
Malaysia Tel. 756.5000/756.5425
Telefax: 755.4424

NETHERLANDS – PAYS-BAS
SDU Uitgeverij
Christoffel Plantijnstraat 2
Postbus 20014
2500 EA's-Gravenhage Tel. (070 3) 78.99.11
Voor bestellingen: Tel. (070 3) 78.98.80
Telefax: (070 3) 47.63.51

NEW ZEALAND
NOUVELLE-ZÉLANDE
Legislation Services
P.O. Box 12418
Thorndon, Wellington Tel. (04) 496.5652
Telefax: (04) 496.5698

NORWAY – NORVÈGE
Narvesen Info Center - NIC
Bertrand Narvesens vei 2
P.O. Box 6125 Etterstad
0602 Oslo 6 Tel. (02) 57.33.00
Telefax: (02) 68.19.01

PAKISTAN
Mirza Book Agency
65 Shahrah Quaid-E-Azam
Lahore 3 Tel. 66.839
Telex: 44886 UBL PK. Attn: MIRZA BK

PORTUGAL
Livraria Portugal
Rua do Carmo 70-74
Apart. 2681
1117 Lisboa Codex Tel.: (01) 347.49.82/3/4/5
Telefax: (01) 347.02.64

BEST COPY AVAILABLE

SINGAPORE – SINGAPOUR
Information Publications Pte
Golden Wheel Bldg.
41, Kallang Pudding, #04-03
Singapore 1334 Tel. 741.5166
Telefax: 742.9356

SPAIN – ESPAGNE
Mundi-Prensa Libros S.A.
Castelló 37, Apartado 1223
Madrid 28001 Tel. (91) 431.33.99
Telefax: (91) 575.39.98

Libreria Internacional AEDOS
Consejo de Ciento 391
08009 – Barcelona Tel. (93) 488.34.92
Telefax: (93) 487.76.59
Llibreria de la Generalitat
Palau Moja
Rambla dels Estudis, 118
08002 – Barcelona
(Subscriptions) Tel. (93) 318.80.12
(Publications) Tel. (93) 302.67.23
Telefax: (93) 412.18.54

SRI LANKA
Centre for Policy Research
c/o Colombo Agencies Ltd.
No. 300-304, Galle Road
Colombo 3 Tel. (1) 574240, 573551-2
Telefax: (1) 575394, 510711

SWEDEN – SUÈDE
Fritzes Fackboksföretaget
Box 16356
Regeringsgatan 12
103 27 Stockholm Tel. (08) 23.89.00
Telefax: (08) 20.50.21

Subscription Agency-Agence d'abonnements
Wennergren-Williams AB
Nordenflychtsvägen 74
Box 30004
104 25 Stockholm Tel. (08) 13.67.00
Telefax: (08) 618.62.32

SWITZERLAND – SUISSE
Maditec S.A. (Books and Periodicals - Livres et périodiques)
Chemin des Palettes 4
1020 Renens/Lausanne Tel. (021) 635.08.65
Telefax: (021) 635.07.80

Mail orders only - Commandes par correspondance seulement
Libraire Payot
C.P. 3212
1002 Lausanne Telefax: (021) 311.13.92

Librairie Unilivres
6, rue de Candolle
1205 Genève Tel. (022) 320.26.23
Telefax: (022) 329.73.18

Subscription Agency - Agence d'abonnement
Naville S.A.
38 avenue Vibert
1227 Carouge Tel.: (022) 308.05.56/57
Telefax: (022) 308.05.88

See also - Voir aussi :
OECD Publications and Information Centre
Schedestrasse 7
D-W 5300 Bonn 1 (Germany)
Tel. (49.228) 21.60.45
Telefax: (49.228) 26.11.04

TAIWAN – FORMOSE
Good Faith Worldwide Int'l. Co. Ltd.
9th Floor, No. 118, Sec. 2
Chung Hsiao E. Road
Taipei Tel. (02) 391.7396/391.7397
Telefax: (02) 394.9176

THAILAND – THAILANDE
Suksit Siam Co. Ltd.
113, 115 Fuang Nakhon Rd.
Opp. Wat Rajbopith
Bangkok 10200 Tel. (662) 251.1630
Telefax: (662) 236.7783

TURKEY – TURQUIE
Kültur Yayinlari Is-Türk Ltd. Sti.
Atatürk Bulvari No. 191/Kat. 13
Kavaklidere/Ankara Tel. 428.11.40 Ext. 2458
Dolmabahce Cad. No. 29
Besiktas/Istanbul Tel. 160.71.88
Telex: 43482B

UNITED KINGDOM – ROYAUME-UNI
HMSO
Gen. enquiries Tel. (071) 873 0011
Postal orders only:
P.O. Box 276, London SW8 5DT
Personal Callers HMSO Bookshop
49 High Holborn, London WC1V 6HB
Telefax: (071) 873 8200
Branches at: Belfast, Birmingham, Bristol, Edinburgh, Manchester

UNITED STATES – ÉTATS-UNIS
OECD Publications and Information Centre
2001 L Street N.W., Suite 700
Washington, D.C. 20036-4910 Tel. (202) 785.6323
Telefax: (202) 785.0350

VENEZUELA
Libreria del Este
Avda F. Miranda 52, Aptdo. 60337
Edificio Galipán
Caracas 106 Tel. 951.1705/951.2307/951.1297
Telegram: Libreste Caracas

YUGOSLAVIA – YOUGOSLAVIE
Jugoslovenska Knjiga
Knez Mihajlova 2, P.O. Box 36
Beograd Tel. (011) 621.992
Telefax: (011) 625.970

Orders and inquiries from countries where Distributors have not yet been appointed should be sent to: OECD Publications Service, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Les commandes provenant de pays où l'OCDE n'a pas encore désigné de distributeur devraient être adressées à : OCDE, Service des Publications, 2, rue André-Pascal, 75775 Paris Cedex 16, France.

Subscription to OECD periodicals may also be placed through main subscription agencies.

Les abonnements aux publications périodiques de l'OCDE peuvent être souscrits auprès des principales agences d'abonnement.

OECD PUBLICATIONS, 2 rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(96 92 03 1) ISBN 92-64-13726-2 - No. 46031 1992

BEST COPY AVAILABLE

The OECD International Education Indicators

A Framework for Analysis

At a time when education is one of the first concerns in OECD countries, developing a set of reliable and useful indicators has become essential to policy formulation and monitoring. This publication examines conceptual issues surrounding the functions and limitations of education indicators that could meet the growing demand for more and better information about the quality of education. The aims, assumptions and procedures of the OECD's Project on International Education Indicators (INES) are described and a set of guiding principles for the continuing development of a set of indicators is proposed.

(96 92 03 1) ISBN 92-64-13726-2 FF 90

BEST COPY AVAILABLE